

Hypersala 1970

**THE ONE COMPONENT MODEL
WITH
INCOMPLETE DATA**

A. CHRISTOFFERSSON

**THE ONE COMPONENT MODEL
WITH
INCOMPLETE DATA**

BY

ANDERS CHRISTOFFERSSON

Fil.lic., Norrl.

INAUGURAL DISSERTATION

BY DUE PERMISSION OF THE FACULTY OF SOCIAL SCIENCES
OF THE UNIVERSITY OF UPPSALA
TO BE PUBLICLY DISCUSSED IN LECTURE ROOM IV
ON OCTOBER 2ND, 1970, AT 10 A.M.,
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41548912_0075

From the Institute of Statistics at the University of Uppsala, Uppsala, Sweden
Head: Professor Herman Wold

THE ONE COMPONENT MODEL
WITH INCOMPLETE DATA

by

Anders Christoffersson



CONTENTS

PREFACE	1
INTRODUCTION	3
SECTION	
1. THEORETICAL MODEL	7
2. ESTIMATED MODEL FOR COMPLETE DATA	11
2.1 Relation between the one factor model and the analysis of variance of a two way classification	14
3. ESTIMATION WHEN THE DATA IS INCOMPLETE. THE MO METHOD	16
4. LARGE SAMPLE PROPERTIES OF THE MO METHOD	24
4.1 First order derivatives	27
4.2 Second order derivatives	31
4.3 Consistency	37
4.4 Large sample variances and covariances of the loading in terms of theoretical parameters	37
4.5 Convergence of the iterative procedure	42
4.6 Limiting distribution of the estimated component and residuals	43
4.7 Estimation of large sample variances and covariances of the loadings in terms of estimated parameters	51
4.8 Confidence regions for the loadings	55
4.9 Testing of the model	57
5. EXPLAINED AND UNEXPLAINED VARIATION	58
6. THE ONE COMPONENT MODEL WITH UNKNOWN LOCATION PARAMETERS	70
6.1 Estimation when the data is incomplete	70

7.	MONTE CARLO EXPERIMENTS	75
8.	APPLICATIONS	82
9.	THE MULTIFACTOR (COMPONENT) MODEL	95
REFERENCES		104
APPENDIX: Tables for the Monte Carlo experiments		107

PREFACE

In 1965, the NIPALS (Non-linear Iterative Partial Least Square) method for interdependent systems was established by Wold (Wold, 1965). It was realized at the outset that the NIPALS approach was of a very wide scope and that it could be used to estimate a large variety of models. In order to investigate the possible uses of the approach, especially in areas where earlier estimation methods are unsatisfactory or altogether lacking, a research program (ICKE-LINEÄRA KUNSKAPSMODELLER. NIPALS-estimation jämte tillämpningar) was set forth by Wold in 1965. This research program was (and is still) financially supported by the Tercentenary Fund of the National Bank of Sweden.

In order to explore the properties of a NIPALS procedure for the one component model with missing observations, which I just had established, I was asked by Professor Wold to participate in his research program. The results of this investigation are presented in this monograph.

For introducing me to the NIPALS approach and giving me the opportunity to participate in the research program and for many stimulating discussions and valuable suggestions during the years I have worked on this monograph, I take the opportunity to express my deep gratitude to my teacher, Professor Herman Wold.

My thanks are also extended to my colleagues at the Institute of Statistics, especially to Docent E. Lyttkens, for his interest in my work and the profit that I have gained from many stimulating discussions. Other rewarding discussions were held with: Docent P. Seeger, Docent A. Taube, L. Bodin, FL and A. Ågren, FL. I also wish to thank Professor B. Rosén for valuable suggestions and comments about the derivation of the limiting distributions in section 4.6 of this monograph.

Further, I am very grateful to Mrs A. Harcourt, M.A. and H. Hutt, B.S., for helping me with the English texts and to Miss I. Sjöberg for typing the manuscript.

Finally, I am indebted to the Tercentenary Fund of the Swedish National Bank for financial support.

Uppsala, May 1970

Anders Christoffersson

INTRODUCTION

Missing observations occur in practically all fields of applied statistics. The methods of resolving this problem are many and they vary with the different fields, but there are two distinct classifications of missing observations:

- i) complete nonresponse
- ii) partial nonresponse

Complete nonresponse encompasses those cases in which no measurement whatever is obtained from some of the selected units. This can occur in sample surveys, for example in a mail survey, in which some proportion of the selected individuals never bother to reply. (See for example Dalenius (1957).) The second type, partial nonresponse, occurs in multivariate analysis. This type of missing observation is characterized by the lack of a complete set of measurements of all variables under analysis for at least some of the selected units. This incompleteness may occur by chance or may lie in the intrinsic nature of the observed situation. An illustrative example of this latter kind of missing observations is given by Wold (1966a), where he studies the problem of estimating racing ability for horses in the United States. In this example there are about 40,000 horses which participate in 10,000 races during one season. That is, we have about 40,000 selected units and 10,000 variables. As just ten horses are entered in each race we have a situation where the observational material is extremely incomplete. Complete data is obviously impossible, as not all the horses can participate in all the races.

In multivariate statistics, the problem has had considerable treatment in the literature. Yates (1933) considers the case of regression analysis in which there are missing observations among the independent variables. In balanced designs with covariates, the problem has been studied by Bartlett (1937), and the estimation of the parameters in the multivariate normal distribution has been treated by Lord (1955), Anderson (1957) and others. Most work in this latter field deals with situations where the missing observations follow specific patterns. However, one method for solving the missing value problem which does not require the missing observations to follow any specific pattern has recently been given by Hocking and Smith (1968). A more exten-

sive list of references can be found in Afifi and Elashoff (1966).

In this monograph we shall consider partial nonresponse in the one factor (component) model.¹ To explain clearly what we intend to investigate let us assume that several psychological tests are given to a group of individuals and that the variability in the tests can be explained by a common latent variable. Let us also assume that we do not have observations on every test for some or all of the individuals. In this situation, there have been two main methods for handling the problem arising from missing observations. These are

- i) the missing observations are supplemented by the mean values for the corresponding tests. (See for example Wold (1966 a).)
- ii) all individuals for whom we do not have observations on all tests are excluded.

These methods for treating incomplete data may be satisfactory if the proportion of missing observations is small. However, if, on the whole, the missing observations in one test correspond to large scores for the latent variable and in another test to small scores, the first method may lead to inconsistent estimates. We may also get inconsistencies if the proportions of missing observations are different for the various tests. The second method, on the other hand, in which we exclude individuals for whom the results are incomplete may give no data to analyse or so little in comparison to the number of individuals tested, that the results are meaningless.

Most methods for factor analysis are based on the correlation (covariance) matrix for the tests, it being assumed (either explicitly or implicitly) that there are no missing observations. In the situation we are considering, one possibility is to estimate the correlation between two tests by using only the individuals who have participated in both tests. But there can still remain correlations which cannot be estimated or are based on so few individuals that the interpretation of the results will be open to doubt.

This difficulty can be avoided in part if there are some tests for which we have observations on every individual since we may then employ one of the methods given in Horst (1965) which utilize only part of the correlation matrix. These methods may work satisfactorily even though they do not make optimal use of the available data. We will discuss these methods in

¹ The model is related to the model used for the analysis of variance of a two way classification. This relationship will be discussed in section 2.

section nine in connection with the multifactor model.

The purpose of this monograph is to present a new method, the MO method, for the estimation of the parameters in the one factor (component) model. This new technique is a suitable adaptation of the NIPALS (Non-linear Iterative Partial Least Square) method for principal components (Wold, (1966a). A report of the seminar where the technique was first presented appears in the initial publication by Wold on the NIPALS approach (Wold, 1966a). The method, which is designed for missing observations in the sense of partial nonresponse, is a least squares method and as such makes use of all the data available. If the assumption of an underlying normal distribution is fulfilled, it is also a maximum likelihood method. It can handle certain types of bias in the process that generates the observations and is consistent even if the missing observations for one test correspond to large scores on the latent variable and for another to small scores.

The model will be presented in sections one and two and the MO method in section three. The large sample properties of the MO method are treated in sections four and five. The one factor (component) model with unknown location parameters is discussed in section six. Some illustrations by Monte Carlo experiments are given in section seven and in section eight we present two applications of the MO method. Finally, in section nine we give a brief discussion of the multifactor (component) model.

1. THEORETICAL MODEL

Component and factor analysis have their origin in the works by Pearson (1901) and Spearman (1904) and have had their greatest use in psychology. The statistical basis for these models was formulated both by Hotelling (1933), who developed the principal component method, and by Lawley (1940) who developed the maximum likelihood method. Other important contributions have been made by Bartlett (1950), Whittle (1952), Rao (1955), Anderson and Rubin (1956), Jöreskog (1963), Lyttkens (1966) and Wold (1966b). Until recently, factor analysis and component analysis were regarded as separate ways of tackling the same problem to arrive at a common answer. However, Rao (1955) revealed that factor and component analysis, in fact, treat two different problems, with factor analysis explaining the correlation in the data set and component analysis explaining the variance. The distinguishing characteristics and similarities between factor and component analysis were completely clarified by Lyttkens (1966), who gave an explicit mathematical expression rather than a heuristic formulation. (An extensive discussion of the relations between factor and component analysis is given in Wold (1966b).) Using the results obtained by Lyttkens we write the theoretical model in the following way:

Let the random variables η_j , $j = 1, 2, \dots, p$ have the following representation:

$$\begin{aligned} \eta_1 &= \gamma_1 \xi + \epsilon_1 \\ (1) \quad \eta_2 &= \gamma_2 \xi + \epsilon_2 \\ &\dots\dots\dots \\ \eta_p &= \gamma_p \xi + \epsilon_p \end{aligned}$$

where the constants γ_j , $j = 1, 2, \dots, p$. (usually called loadings), are normalized so that

$$(1a) \quad \sum_j^p \gamma_j^2 = 1$$

and the ϵ_j , $j = 1, 2, \dots, p$ are residuals which have zero expectation and are

uncorrelated with the latent variable ξ . In addition, the second order moments of η_j , $j = 1, 2 \dots p$, are assumed to exist.

Multiplying each equation in (1) with the corresponding γ_j and summing over j , we obtain

$$(1b) \quad \sum_j^p \gamma_j \eta_j = \xi \sum_j^p \gamma_j^2 + \sum_j^p \gamma_j \epsilon_j = \xi + \sum_j^p \gamma_j \epsilon_j$$

Now, this model can be interpreted in two ways:

i) If it is assumed that

$$(1c) \quad \sum_j^p \gamma_j \epsilon_j = 0$$

we arrive at the principal component model, in which the component ξ can be written as

$$(1d) \quad \xi = \sum_j^p \gamma_j \eta_j$$

ii) If the residuals are assumed to be mutually uncorrelated, we obtain the one factor model, where the factor ξ can be written as

$$(1e) \quad \xi = \sum_j^p \gamma_j \eta_j - \sum_j^p \gamma_j \epsilon_j$$

As the residuals are uncorrelated, it follows that in general

$$(1f) \quad \sum_j^p \gamma_j \epsilon_j \neq 0$$

and hence that the factor and the component differ. However, when the residuals have equal variances, the loadings are the same for both models, and in particular, this factor model then becomes a special case of the general factor model. In this monograph we shall assume that the residuals have equal variances, whereas the general factor model assumes only that the residuals are uncorrelated. (See for example Jöreskog (1963).)

We shall rewrite the models so as to emphasize the difference between the factor representation and the component representation. The factor representation of the model (1) shall be written:

$$(2) \quad \eta_j = \gamma_j \mu + \epsilon_j \quad j = 1, 2, \dots, p$$

with η_j as the j :th observed variable,

μ as the latent factor,

γ_j as the j :th loading, normalized so that

$$\sum_j^p \gamma_j^2 = 1,$$

and ϵ_j as a stochastic residual uncorrelated with ϵ_s for every $s \neq j$, with expectation equal to zero and constant variance equal to σ^2 for every j .

Furthermore, the residuals are independent of the factor.

The component representation of the model (1) shall be written

$$(3) \quad \eta_j = \gamma_j \xi + \delta_j \quad j = 1, 2, \dots, p$$

where the observed variables η_j and the loadings γ_j are the same as in (2). We define the component ξ as

$$(3a) \quad \xi = \mu + \sum_j^p \gamma_j \epsilon_j$$

and the residual δ_j as

$$(3b) \quad \delta_j = \epsilon_j - \gamma_j \sum_s^p \gamma_s \epsilon_s \quad j = 1, 2, \dots, p$$

It is easily verified that

$$(3c) \quad \xi = \sum_j^p \gamma_j \eta_j$$

and

$$(3d) \quad \sum_j^p \gamma_j \delta_j = 0$$

and that the component, ξ , is uncorrelated with the residual δ_j for $j = 1, 2, \dots, p$.

If we let σ_μ^2 denote the variance of the factor, it follows easily from (3a) that the variance of the component, σ_ξ^2 , is given by

$$(3e) \quad \sigma_\xi^2 = \sigma_\mu^2 + \sigma^2$$

In applications of these models, the main interest focuses on estimation of the loadings, on the variance explained by the factor or by the component, and on the estimation of the factor or of the component itself.

2. ESTIMATED MODEL FOR COMPLETE DATA

Our observations can be arranged in a matrix $Y = (y_{ij})$ where y_{ij} denotes the i :th observation of η_j . Let μ_i denote the factor score corresponding to the i :th observation and ϵ_{ij} the residual corresponding to y_{ij} . The theoretical model for our observations can then be written as

$$(4) \quad y_{ij} = \gamma_j \mu_i + \epsilon_{ij} \quad j = 1, 2, \dots, p; i = 1, 2, \dots, k$$

where k is the number of observations. In our example cited earlier about psychological tests, y_{ij} would refer to the score in the j :th test for the i :th person, μ_i to the factor score for the i :th person and ϵ_{ij} to the residual.

In the remainder of the monograph we shall use the convention, provided that it is clear from the context, that subscripts shall range over their respective values.

For the component representation of the model, we have

$$(5) \quad y_{ij} = \gamma_j \xi_i + \delta_{ij}$$

with the following equations corresponding to (3a-d)

$$(5a) \quad \xi_i = \mu_i + \sum_j^p \gamma_j \epsilon_{ij}$$

$$(5b) \quad \delta_{ij} = \epsilon_{ij} - \gamma_j \sum_s^p \gamma_s \epsilon_{is}$$

$$(5c) \quad \xi_i = \sum_j^p \gamma_j y_{ij}$$

$$(5d) \quad \sum_j^p \gamma_j \delta_{ij} = 0$$

If we use the method of least squares or assume normal distribution in (4) and (5) and use the method of maximum likelihood, we find that, for the

same set of data, the two models give identical estimates of the loadings and that the estimate of the factor in (4) and the component in (5) are identical. Thus neither the method of least squares nor the method of maximum likelihood discriminates between the two models.

Actually, both methods yield the sample component as the estimate of the factor, but since (as will be shown in section 4.6) the sample component tends to the theoretical component when the sample size (k) increases, we conclude that the estimate of the factor given by either method is inconsistent. Therefore we may write down the same estimated model for both the factor representation and the component representation, as follows

$$(6) \quad y_{ij} = g_j x_i + e_{ij}$$

where g_j , the estimate of the loading γ_j , is normalized so that

$$(6a) \quad \sum_j^p g_j^2 = 1$$

and x_i is the estimate of the component score ξ_i or of the factor score μ_i . The residual in the sample is denoted by e_{ij} .

For the case of complete data, the maximum likelihood solution has been given by Young (1941) and the least square solution by Whittle (1952). Wold (1966a) has derived the following iterative procedure for the estimation of g_j and x_i .

Given the s :th approximation of the loadings, $g_j^{(s)}$, where

$$\sum_j^p \left(g_j^{(s)}\right)^2 = 1$$

we calculate the $(s+1)$:st approximation of x_i as

$$(7a) \quad x_i^{(s+1)} = \sum_j^p y_{ij} g_j^{(s)}$$

and the $(s+1)$:st approximation of the loadings as

$$(7b) \quad G_j^{(s+1)} = \sum_i^k y_{ij} x_i^{(s+1)}$$

and then,

$$(7c) \quad g_j^{(s+1)} = \frac{G_j^{(s+1)}}{\left[\sum_j^p \left(G_j^{(s+1)} \right)^2 \right]^{1/2}}$$

In the limit we have

$$(7d) \quad g_j = \lim_{s \rightarrow \infty} g_j^{(s)}; \quad x_i = \lim_{s \rightarrow \infty} x_i^{(s)}$$

The above method (7a-c) gives the same estimates as the method of least squares, and therefore minimizes the sum of squared residuals:

$$(8) \quad Q = \sum_i^k \sum_j^p (y_{ij} - g_j x_i)^2$$

Further, it follows from eq:s 7a-d that

$$(8a) \quad \sum_i^k e_{ij} x_i = 0$$

$$(8b) \quad \sum_j^p e_{ij} g_j = 0$$

$$(8c) \quad x_i = \sum_j^p y_{ij} g_j$$

Comparing (8a-c) with the theoretical specification of the model, we find that (8a) is the analogue of the zero correlation between the component and the residuals while (8b) corresponds to (1c) and (8c) to (1d).

2.1 Relation between the one factor model and the analysis of variance of a two way classification.

As mentioned earlier, the factor model (4) is closely related to the models used in the analysis of variance of a two way classification. The factor model was used by Fisher and Mackenzie (1923) when analysing the manurial response of different potato varieties. In their application, the yield was assumed to be a product of two "factors", one depending on the variety and the other on the manure. The unknowns in the model were estimated by least squares, i.e. by minimizing (8). An approximate solution was obtained by iterating 1 1/2 cycles according to a process similar to (7a-c).

The general model used in variance analysis of two-way classified data can be written

$$(9) \quad y_{ij} = \alpha_j + \beta_i + (\alpha\beta)_{ij} + \epsilon_{ij}$$

where α_j corresponds to column effects, β_i to row effects, $(\alpha\beta)_{ij}$ to interaction between columns and rows and ϵ_{ij} is the residual. If there is only one observation per cell (in the classical terminology of variance analysis: one observation for each combination of treatment and variety) it is not possible to test the hypothesis that there are no interactions. It is then natural to try to impose some restrictions on the form of the interaction. One reasonable assumption (Mandel, 1961) is that the interactions are proportional to the row (or column) effects. The model (9) then takes the form

$$(9a) \quad y_{ij} = \alpha_j + \beta_i + \gamma_j \beta_i + \epsilon_{ij} = \alpha_j + \gamma_j^{\star} \beta_i + \epsilon_{ij}$$

This is essentially the same as (4). The two models will be identical if in (4) we let y_{ij} denote the deviation from the corresponding column mean.

To estimate the parameters of the model, Mandel (1961) applies a two step procedure. To begin with, the α_j and β_i are estimated as if there were no interactions and then the $\gamma_j^{\star}$ are estimated, using the estimated values of the β_i . Mandel also shows how tests of significance should be carried out given that the parameters have been estimated by the above method.

If the parameters of the model (9a) are estimated by maximum likelihood

(assuming that the residuals are independently normally distributed) the estimates of the α_j are the means of the corresponding columns. The estimates of $\gamma_j^\star$ and β_i can be obtained from (7a-c) by putting the deviation from the sample mean of the corresponding column equal to y_{ij} .

In the model (9a), the hypothesis that there are no row effects implies that the factor has zero variance, and the hypothesis of no interactions implies that the loadings are equal. The stochastic properties of the maximum likelihood estimates have been derived by Anderson (1963). Using these results the appropriate hypotheses may be tested.

The model (9a), can be generalized in the following way

$$(9b) \quad y_{ij} = \alpha_j + \beta_i + \gamma_j \mu_i + \epsilon_{ij}$$

i.e. the interactions are not necessarily proportional to the row (or column) effects.

If we assume that the residuals are independently normally distributed one set of maximum likelihood estimates is

$$(9c) \quad \hat{\alpha}_j = \frac{1}{k} \sum_i^k y_{ij} \quad \text{as the estimate of } \alpha_j$$

$$(9d) \quad \hat{\beta}_i = \frac{1}{p} \sum_j^p (y_{ij} - \hat{\alpha}_j) \quad \text{as the estimate of } \beta_i$$

The estimates of γ_j and μ_i are given by the process (7a-c) in which $(y_{ij} - \hat{\alpha}_j - \hat{\beta}_i)$ is substituted for y_{ij} . The maximum likelihood solution is unique up to an additive constant for $\{\hat{\alpha}_j, \hat{\beta}_i\}$ and up to a multiplicative constant for the estimates of γ_j and μ_i . For example, adding a constant to the $\hat{\alpha}_j$'s and subtracting the same constant from the $\hat{\beta}_i$'s gives the same likelihood.

This model (9b) can be further generalized by letting the interactions be written as a sum of m products, i.e. $(\alpha\beta)_{ij} = \sum_r^m \gamma_{jr} \mu_{ir}$. For a more detailed discussion of this model reference is made to Gollob (1968).

3. ESTIMATION WHEN DATA IS INCOMPLETE. THE MO METHOD

As mentioned in the introduction, the problem with missing observations has been treated mainly in two ways:

- i) Estimation is made by using the mean value over the j .th variable (Wold, 1966a) as a substitute for missing values.
- ii) We exclude all individuals who have not participated in each test.

Both these ways of treating incomplete data may be satisfactory if the proportion of missing observations is small. But, as stated before, if this is not the case, (i) may lead to inconsistent estimates if the missing observations in one variable correspond either to large or small factor scores. Inconsistencies may also occur when the proportions of missing observations are different for different tests. This latter case is illustrated by a small Monte Carlo experiment in section seven. An example of the limited scope of the second approach is given in section eight.

In this section we present a new method, the MO method, to handle incomplete data. First, we must rewrite the component representation of the model by introducing the indicator variable c_{ij} , which is defined by

$$(10) \quad c_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is observed} \\ 0 & \text{if } y_{ij} \text{ is not observed} \end{cases}$$

By assuming that

$$(11) \quad \sum_j^p c_{ij} \gamma_j^2 \neq 0$$

we can then define the component as

$$(12) \quad \xi_i = \frac{\sum_j^p c_{ij} \gamma_j y_{ij}}{\sum_j^p c_{ij} \gamma_j^2}$$

and, as in (5a) we obtain

$$(12a) \quad \xi_i = \mu_i + \frac{\sum_j^p c_{ij} \gamma_j \epsilon_{ij}}{\sum_j^p c_{ij} \gamma_j^2}$$

Thus, the component representation of the model can be written as

$$(13) \quad y_{ij} = \gamma_j \xi_i + \delta_{ij}$$

where

$$(13a) \quad \delta_{ij} = \epsilon_{ij} - \gamma_j \frac{\sum_j^p c_{ij} \gamma_j \epsilon_{ij}}{\sum_j^p c_{ij} \gamma_j^2}$$

It is then easily verified that

$$(13b) \quad \sum_j^p c_{ij} \gamma_j \delta_{ij} = 0$$

Comment

This way of specifying the component representation of the model can be obtained from (5a-d) by supplementing the original data with the corresponding systematic parts of the missing observations. That is, if y_{ij} is missing we put $\gamma_j \mu_i$ in its place.

The estimated model is written as in the preceeding section, i.e.

$$(14) \quad y_{ij} = g_j x_i + e_{ij}$$

where $e_{ij} = 0$ if $c_{ij} = 0$.

Using (8a-b) and the normalising condition (6a) we find:

$$(14a) \quad \sum_i^k e_{ij} x_i = \sum_i^k c_{ij} (y_{ij} - g_j x_i) x_i = 0$$

$$(14b) \quad \sum_j^p e_{ij} g_j = \sum_j^p c_{ij} (y_{ij} - g_j x_i) g_j = 0$$

$$(14c) \quad \sum_j^p g_j^2 = 1$$

These expressions can be rewritten to give

$$(15a) \quad g_j = \frac{\sum_i^k c_{ij} y_{ij} x_i}{\sum_i^k c_{ij} x_i^2} \quad \text{if } \sum_i^k c_{ij} x_i^2 \neq 0$$

$$(15b) \quad x_i = \frac{\sum_j^p c_{ij} y_{ij} g_j}{\sum_j^p c_{ij} g_j^2} \quad \text{if } \sum_j^p c_{ij} g_j^2 \neq 0$$

$$(15c) \quad \sum_j^p g_j^2 = 1$$

(Unless otherwise stated it will be assumed, throughout the monograph, that

$\sum_i^k c_{ij} x_i^2 \neq 0$ and $\sum_j^p c_{ij} g_j^2 \neq 0$ for every i and j .)

The MO estimator is defined to be the solution of (15a-c). Equations (15a-c) define the estimates g_j and x_i as the solutions of a nonlinear system, but because of the complex nonlinearity there is no simple explicit solution. In effect, (15a-c) define the loadings as an eigenvector to some nonlinear operator. In the case of complete data it is readily seen that the system (15a-c) reduces to the linear case:

$$(16a) \quad g_j = \frac{\sum_i^k y_{ij} x_i}{\sum_i^k x_i^2}$$

$$(16b) \quad x_i = \frac{\sum_j^p y_{ij} g_j}{\sum_j^p g_j^2}$$

$$(16c) \quad \sum_j^p g_j^2 = 1$$

Eqs (16a-c) define the loadings as an eigenvector to a linear operator and it is this system that is solved by the iterative procedure (7a-c). The above system has, in general, p orthogonal solutions, each of which is associated with an eigenvalue of the operator defined in (16a-c). The solution corresponding to the largest eigenvalue gives a minimum for the sum of the squared residuals (8). So, in this case, there is no difficulty in choosing between the solutions. In addition, the iterative process (7a-c) will converge to the eigenvector that corresponds to the largest eigenvalue, provided the starting point of the iteration is not exactly orthogonal to this eigenvector (Lyttkens, 1966). In practice, the procedure always converges to the correct solution.

For the system (15a-c) the situation is somewhat different, as we do not know the exact number of solutions. It is not even possible to show that the system always has nondegenerate solutions, for one can construct so called "pathological" data sets for which the denominator in (15b) will tend to zero, and thus produce only degenerate solutions. Fortunately, this phenomenon rarely occurs. (It has happened only twice in over 10,000 test runs on the computer.)

To solve the system (15a-c) the following iterative procedure is proposed:

Having obtained the s :th approximation of the loadings, $g_j^{(s)}$, the s :th approximation of the component $x_i^{(s)}$ is defined by

$$(17a) \quad x_i^{(s)} = \frac{\sum_j^p c_{ij} y_{ij} g_j^{(s)}}{\sum_j^p c_{ij} \left(g_j^{(s)}\right)^2}$$

The $(s+1)$:st approximation of the loadings is derived in two stages. We first obtain

$$(17b) \quad G_j^{(s+1)} = \frac{\sum_i^k c_{ij} y_{ij} x_i^{(s)}}{\sum_i^k c_{ij} \left(x_i^{(s)}\right)^2}$$

and then obtain

$$(17c) \quad g_j^{(s+1)} = \frac{G_j^{(s+1)}}{\left[\sum_j^p \left(G_j^{(s+1)}\right)^2\right]^{1/2}}$$

In practical applications this approach works sufficiently well. Non convergence has occurred in one application in which over 80% of the observations were missing. (Christoffersson, 1970a.) The behaviour of the procedure for large samples will be treated later in the monograph.

The MO estimates can also be derived by least squares. The sum of the squared residuals becomes

$$(18) \quad q = \sum_i^k \sum_j^p c_{ij} (y_{ij} - g_j x_i)^2$$

Using the method of Lagrange, we minimize q subject to the condition

$$\sum_j^p g_j^2 = 1:$$

$$(18a) \quad Q = q + \lambda \left(\sum_j^p g_j^2 - 1 \right)$$

$$= \sum_i^k \sum_j^p c_{ij} (y_{ij} - g_j x_i)^2 + \lambda \left(\sum_j^p g_j^2 - 1 \right)$$

The partial derivatives of Q with respect to g_j and x_i are:

$$(18b) \quad \frac{\delta Q}{\delta g_j} = 2 \sum_i^k c_{ij} (y_{ij} - g_j x_i) (-x_i) + 2\lambda g_j$$

$$(18c) \quad \frac{\delta Q}{\delta x_i} = 2 \sum_j^p c_{ij} (y_{ij} - g_j x_i) (-g_j)$$

Setting the partial derivatives equal to zero we obtain

$$(19a) \quad g_j = \frac{\sum_i^k c_{ij} y_{ij} x_i}{\sum_i^k c_{ij} x_i^2 + 2\lambda}$$

$$(19b) \quad x_i = \frac{\sum_j^p c_{ij} y_{ij} g_j}{\sum_j^p c_{ij} g_j^2}$$

To show that the multiplier λ is zero we may proceed as follows:

Define:
$$\begin{cases} y_{ij}^\star = y_{ij} & \text{if } c_{ij} = 1 \\ y_{ij}^\star = g_j x_i & \text{if } c_{ij} = 0 \end{cases}$$

Using the normalizing condition on the loadings, we can then rewrite eq:s (19a-b) as follows

$$(20a) \quad g_j = \frac{\sum_i^k y_{ij}^\star x_i}{\sum_i^k x_i^2 + 2\lambda}$$

$$(20b) \quad x_i = \frac{\sum_j^p y_{ij}^{\star} g_j}{\sum_j^p g_j^2} = \sum_j^p y_{ij}^{\star} g_j$$

Multiplying (20a) by g_j and summing over j gives

$$\begin{aligned} \sum_j^p g_j^2 &= \frac{1}{\sum_i^k x_i^2 + 2\lambda} \sum_j^p \sum_i^k y_{ij}^{\star} x_i g_j \\ &= \frac{1}{\sum_i^k x_i^2 + 2\lambda} \sum_i^k x_i \sum_j^p y_{ij}^{\star} g_j = \frac{\sum_i^k x_i^2}{\sum_i^k x_i^2 + 2\lambda} \end{aligned}$$

But as $\sum_j^p g_j^2 = 1$ it follows that $\lambda = 0$.

The least squares estimates are then defined by

$$(21a) \quad g_j = \frac{\sum_i^k c_{ij} y_{ij} x_i}{\sum_i^k c_{ij} x_i^2}$$

$$(21b) \quad x_i = \frac{\sum_j^p c_{ij} y_{ij} g_j}{\sum_j^p c_{ij} g_j^2}$$

$$(21c) \quad \sum_j^p g_j^2 = 1$$

which are identical to (15a-c).

If the residuals are normally distributed, the logarithm of the likelihood of the sample, given the variance of the residuals, is proportional to the quantity q defined in (18). In such a case, it follows that the MO estimates are maximum likelihood estimates.

4. LARGE SAMPLE PROPERTIES OF THE MO METHOD

If we assume that the residuals in the one factor model (4) are normally distributed, the MO estimates are maximum likelihood estimates.

However, we cannot rely on the large sample properties of maximum likelihood estimates in this case, as the number of unknown parameters increases with the size of the sample, i.e. the number of unknown parameters increases with k . Furthermore, this implies that we cannot expect to obtain consistent estimates of the factor scores. As the number of tests is regarded as constant, it is possible, however, to show that the MO estimates of the loadings are consistent and to obtain the second order moments for the loadings. It is also possible to give a local criterion for the convergence of the iterative procedure defined by (17a-c).

We first show that if we regard the sum of squared residuals (18) divided by k as a function of the loadings only, we can obtain the MO estimates by minimizing this sum. Consider the function f defined by:

$$(22) \quad f = \frac{1}{k} \sum_i^k \sum_j^p c_{ij} (y_{ij} - g_j x_i)^2$$

with the conditions

$$(22a) \quad x_i = \frac{\sum_j^p c_{ij} y_{ij} g_j}{\sum_j^p c_{ij} g_j^2}$$

$$(22b) \quad \sum_j^p g_j^2 = 1$$

By (22a), we can regard the component scores as functions of the loadings. Instead of minimizing f directly, we use the Lagrangian method and thus find the minimum of

$$(23) \quad F = \frac{1}{k} \sum_i^k \sum_j^p c_{ij} (y_{ij} - g_j x_i)^2 + \lambda (\sum_j^p g_j^2 - 1)$$

As we do not use any multipliers to take account of the conditions (22a), we must introduce the following two types of partial derivatives:

- i) $\frac{dF}{\delta g_r}$; wich denotes the partial derivative of F with respect to g_r when F is regarded as a function of $g_1, g_2 \dots g_p$.
- ii) $\frac{\delta F}{\delta g_r}$; which denotes the partial derivative of F with respect to g_r when F is regarded as a function of $g_1, g_2 \dots g_p$ and $x_1, x_2 \dots x_k$.

The partial derivatees of F are:

$$(24) \quad \frac{dF}{\delta g_r} = \frac{\delta F}{\delta g_r} + \sum_i^k \frac{\delta F}{\delta x_i} \frac{\delta x_i}{\delta g_r} = \frac{2}{k} \sum_i^k c_{ir} (y_{ir} - g_r x_i) (-x_i) +$$

$$+ 2\lambda g_r + \sum_i^k 2 \left[\sum_j c_{ij} (y_{ij} - g_j x_i) (-g_j) \right] \frac{\delta x_i}{\delta g_r}$$

where

$$(24a) \quad \frac{\delta x_i}{\delta g_r} = \frac{\sum_j^p c_{ij} g_j^2 c_{ir} y_{ir} - \sum_j^p c_{ij} y_{ij} g_j 2c_{ir} g_r}{\left(\sum_j^p c_{ij} g_j^2 \right)^2}$$

Using (22a) we then have

$$(24b) \quad \frac{\delta x_i}{\delta g_r} = \frac{c_{ir} y_{ir} - 2x_i c_{ir} g_r}{\sum_j^p c_{ij} g_j^2}$$

And as

$$\sum_j^p c_{ij} (y_{ij} - g_j x_i) (-g_j) = 0,$$

we get

$$(25) \quad \frac{dF}{\delta g_r} = \frac{2}{k} \sum_i^k c_{ir} (y_{ir} - g_r x_i) (-x_i) + 2\lambda g_r$$

Comparing (25) and (22a, b) with (18b-c), we conclude that the multiplier is zero, and that the MO estimates can be obtained by minimizing the function f regarded as a function of the loadings.

In order to study the large sample properties of the MO method we must study the partial derivatives of F with respect to the loadings. We shall, in this case, use the factor representation of the model and assume the residuals to be normally distributed. The model under these conditions is

$$(26) \quad y_{ij} = \gamma_j \mu_i + \epsilon_{ij}$$

where

$$(26a) \quad \mu_i \quad \text{are the stochastic or non-stochastic latent factor scores, and } \mu_i \text{ is independent of } \mu_s \text{ for } i \neq s. \text{ We also assume that the expectation and variance of } \mu_i \text{ exist.}$$

$$(26b) \quad \gamma_j \quad \text{are the loadings, normalized so that}$$

$$\sum_j^p \gamma_j^2 = 1$$

and

$$(26c) \quad \epsilon_{ij} \quad \text{is the } ij\text{:th residual which is normally distributed with zero expectation, constant variance, independent of } \mu_s \text{ for every } s, \text{ and } \epsilon_{ij} \text{ is independent of } \epsilon_{rs} \text{ for } r \neq i \text{ and/or } s \neq j. \text{ The variance of } \epsilon_{ij} \text{ is denoted by } \sigma^2.$$

We will further assume that the positions of the missing observations are given and are constant through repeated sampling. (This means that c_{ij} is regarded as nonstochastic.)

Comment

It should be noted that in this specification of the model we allow the factor to have different expectation for different values of i . Furthermore, the missing observations may in one test correspond to large factor scores and in another test to small scores.

4.1 First order derivatives

As the Lagrangian multiplier in (25) is zero, the first order derivatives become

$$(27) \quad \frac{dF}{\delta g_r} = \frac{-2}{k} \sum_i^k c_{ir} (y_{ir} - g_r x_i) x_i$$

Later on, we shall need the expectations, variances and covariances of these derivatives, given that the estimated loadings equal the true loadings.

Before evaluating these expressions, we introduce some convenient notation.

Let

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{pmatrix} \quad \text{and} \quad g = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_p \end{pmatrix}$$

denote the true and estimated loadings written as column vectors. Further, let the symbol E represent the mathematical expectation and the symbol D^2 , the variance defined by $D^2(x) = E \left[[x - E(x)]^2 \right]$

i) Expectation of first order derivatives

We shall show that

$$(28) \quad E \left[\left. \frac{dF}{\delta g_r} \right|_{g=\gamma} \right] = 0$$

Given that $g = \gamma$, it follows that

$$\begin{aligned}
 (29) \quad x_i &= \frac{\sum_j^p c_{ij} y_{ij} \gamma_j}{\sum_j^p c_{ij} \gamma_j^2} = \frac{\sum_j^p c_{ij} \gamma_j^2 \mu_i + \sum_j^p c_{ij} \gamma_j \epsilon_{ij}}{\sum_j^p c_{ij} \gamma_j^2} \\
 &= \mu_i + K_i \sum_j^p c_{ij} \gamma_j \epsilon_{ij} = \mu_i + \bar{\epsilon}_i
 \end{aligned}$$

where

$$(29a) \quad K_i = \frac{1}{\sum_j^p c_{ij} \gamma_j^2}$$

Hence

$$\begin{aligned}
 (30) \quad \left[\frac{dF}{dg_r} \right]_{g=\gamma} &= \frac{-2}{k} \sum_i^k c_{ir} (y_{ir} - \gamma_r \mu_i - \gamma_r \bar{\epsilon}_i) x_i \\
 &= \frac{-2}{k} \sum_i^k c_{ir} (\gamma_r \mu_i + \epsilon_{ir} - \gamma_r \mu_i - \gamma_r \bar{\epsilon}_i) (\mu_i + \bar{\epsilon}_i) \\
 &= \frac{-2}{k} \sum_i^k c_{ir} (\epsilon_{ir} \mu_i + \epsilon_{ir} \bar{\epsilon}_i - \gamma_r \bar{\epsilon}_i \mu_i - \gamma_r \bar{\epsilon}_i \bar{\epsilon}_i)
 \end{aligned}$$

Now

$$\begin{aligned}
 (31a) \quad E(\bar{\epsilon}_i^2) &= E(K_i^2 (\sum_j^p c_{ij} \gamma_j \epsilon_{ij})^2) \\
 &= K_i^2 E(\sum_j^p c_{ij} \gamma_j^2 \epsilon_{ij}^2 + 2 \sum_{s < t}^p c_{is} c_{it} \gamma_s \gamma_t \epsilon_{is} \epsilon_{it}) \\
 &= K_i^2 \sum_j^p c_{ij} \gamma_j^2 \sigma^2 = K_i \sigma^2
 \end{aligned}$$

$$(31b) \quad E(\epsilon_{ir} \bar{\epsilon}_i) = E(K_i \epsilon_{ir} \sum_j^p c_{ij} \gamma_j \epsilon_{ij}) = K_i c_{ir} \gamma_r \sigma^2$$

$$(31c) \quad E(\epsilon_{ir} \mu_i) = 0$$

$$(31d) \quad E(\gamma_r \bar{\epsilon}_i \mu_i) = 0$$

and

$$(31e) \quad E(\bar{\epsilon}_i) = 0, \text{ giving our desired result :}$$

$$E \left[\frac{dF}{dg_r} \Big|_{g=\gamma} \right] = 0$$

That is, given $g = \gamma$, the expectation of the first order derivatives is zero.

ii) Variances and covariances of first order derivatives

We shall also need the variances and covariances of $\frac{dF}{dg_r}$ given that $g = \gamma$. First, we have

$$\begin{aligned}
 (32a) \quad & D^2 [(y_{ir} - \gamma_r x_i) x_i] \\
 &= E [(y_{ir} - \gamma_r x_i)^2 x_i^2] - E^2 [(y_{ir} - \gamma_r x_i) x_i] \\
 &= E [(y_{ir} - \gamma_r x_i)^2] \cdot E(x_i^2) \\
 &= E [(\epsilon_{ir} - \gamma_r \bar{\epsilon}_i)^2] \cdot E[(\mu_i + \bar{\epsilon}_i)^2] \\
 &= E(\epsilon_{ir}^2 - 2\gamma_r \epsilon_{ir} \bar{\epsilon}_i + \gamma_r^2 \bar{\epsilon}_i^2) \cdot E(\mu_i^2 + 2\mu_i \bar{\epsilon}_i + \bar{\epsilon}_i^2)
 \end{aligned}$$

$$\begin{aligned}
&= (\sigma^2 - 2\gamma_r^2 \sigma^2 K_i + \gamma_r^2 \sigma^2 K_i)(\bar{\mu}_i^2 + K_i \sigma^2) \\
&= \sigma^2 (1 - \gamma_r^2 K_i)(\bar{\mu}_i^2 + \sigma^2 K_i)
\end{aligned}$$

where

$$E(\mu_i^2) = \bar{\mu}_i^2$$

(This follows from x_i and $(y_{ir} - \gamma_r x_i)$ having a joint normal distribution and being uncorrelated, and thus independent.)

Secondly,

$$\begin{aligned}
(32b) \quad &E[(y_{ir} - \gamma_r x_i) x_i (y_{is} - \gamma_s x_i) x_i] \\
&= E[(\epsilon_{ir} - \gamma_r \bar{\epsilon}_i)(\epsilon_{is} - \gamma_s \bar{\epsilon}_i) x_i^2] \\
&= E[(\epsilon_{ir} - \gamma_r \bar{\epsilon}_i)(\epsilon_{is} - \gamma_s \bar{\epsilon}_i)] E(x_i^2) \\
&= -\gamma_r \gamma_s K_i \sigma^2 (\bar{\mu}_i^2 + K_i \sigma^2)
\end{aligned}$$

Define

$$(33a) \quad \left[\frac{dF}{d\mathbf{g}_r} \right]_{\mathbf{g}=\boldsymbol{\gamma}} = \mathbf{y}_r$$

and

$$(33b) \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}$$

then

$$(34a) \quad D^2(y_r) = \frac{4\sigma^2}{k^2} \sum_i^k c_{ir} (1 - K_i \gamma_r^2) (\bar{\mu}_i^2 + K_i \sigma^2) = \frac{1}{k} \varphi_{rr}$$

and

$$(34b) \quad \text{cov}(y_r y_s) = \frac{-4\sigma^2}{k^2} \sum_i^k c_{ir} c_{is} \gamma_r \gamma_s K_i (\bar{\mu}_i^2 + K_i \sigma^2) = \frac{1}{k} \varphi_{rs}$$

Comment

φ_{rs} and φ_{rr} are arithmetic means of nonstochastic elements which are bounded above. We will assume that the second order moments of y are stationary in the sense that

$$(35a) \quad \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=t}^{k+t} c_{ir} (1 - K_i \gamma_r^2) (\bar{\mu}_i^2 + K_i \sigma^2)$$

and

$$(35b) \quad \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=t}^{k+t} c_{ir} c_{is} \gamma_r \gamma_s K_i (\bar{\mu}_i^2 + K_i \sigma^2)$$

exist and are constant for all values of t . The asymptotic distribution of $\sqrt{k} y$ will then be normal with covariance matrix $\Phi = (\varphi_{rs})$ and y will tend stochastically to the zero vector.

4.2. Second order derivatives

To study the large sample properties we also need the second order partial derivatives:

$$\begin{aligned} (36a) \quad \frac{d^2 F}{\delta g_r \delta g_s} &= \frac{2}{k} \sum_i^k c_{ir} \left(-y_{ir} \frac{\delta x_i}{\delta g_s} + 2 g_r x_i \frac{\delta x_i}{\delta g_s} \right) \\ &= \frac{2}{k} \sum_i^k c_{ir} (-y_{ir} + 2 g_r x_i) \frac{\delta x_i}{\delta g_s} \\ &= \frac{2}{k} \sum_i^k c_{ir} (-y_{ir} + 2 g_r x_i) \frac{c_{is} (y_{is} - 2 x_i g_s)}{\sum_j^p c_{ij} g_j^2} \end{aligned}$$

$$= \frac{2}{k} \sum_i^k \frac{c_{ir} c_{is}}{\sum_j^p c_{ij} g_j^2} (- y_{ir} y_{is} + 2x_i g_r y_{is} +$$

$$+ 2x_i g_s y_{ir} - 4g_r g_s x_i^2)$$

and

$$(36b) \quad \frac{d^2 F}{\delta g_r^2} = \frac{2}{k} \sum_i^k c_{ir} \left(- y_{ir} \frac{\delta x_i}{\delta g_r} + x_i^2 + 2g_r x_i \frac{\delta x_i}{\delta g_r} \right)$$

$$= \frac{2}{k} \sum_i^k c_{ir} \left(x_i^2 + (-y_{ir} + 2g_r x_i) \frac{c_{ir} (y_{ir} - 2x_i g_r)}{\sum_j^p c_{ij} g_j^2} \right)$$

$$= \frac{2}{k} \sum_i^k c_{ir} x_i^2 + \sum_i^k \frac{c_{ir}}{\sum_j^p c_{ij} g_j^2} (-y_{ir}^2 + 4x_i g_r y_{ir} - 4g_r^2 x_i^2)$$

i) Expectation of second order derivatives

From (29) it is easily seen that

$$(37a) \quad y_{ij} = \gamma_j \mu_i + \epsilon_{ij} = \gamma_j x_i + \epsilon_{ij} - \gamma_j \bar{\epsilon}_i = \gamma_j x_i + \delta_{ij}$$

In section 4.1, we have shown that

$$(37b) \quad E(x_i \delta_{ij}) = E [(\mu_i + \bar{\epsilon}_i) (\epsilon_{ij} - \gamma_j \bar{\epsilon}_i)] = 0$$

From (29) and (37a) it follows that

$$(38) \quad \left[\frac{d^2 F}{\delta g_r \delta g_s} \Big|_{g=\gamma} \right] = \frac{2}{k} \sum_i^k c_{ir} c_{is} K_i [-y_{ir} y_{is} + \\ + 2x_i (\gamma_s \delta_{ir} + \gamma_r \delta_{is})]$$

and that its expectation can thus be expressed as:

$$(39a) \quad E \left[\frac{d^2 F}{\delta g_r \delta g_s} \Big|_{g=\gamma} \right] = \frac{2}{k} \sum_i^k K_i c_{ir} c_{is} (-E(y_{ir} y_{is}) + \\ + 2E [(x_i) (\gamma_s \delta_{ir} + \gamma_r \delta_{is})]) \\ = \frac{2}{k} \sum_i^k K_i c_{ir} c_{is} (-\gamma_r \gamma_s \bar{\mu}_i^2) \\ = \frac{-2 \gamma_r \gamma_s}{k} \sum_i^k c_{ir} c_{is} K_i \bar{\mu}_i^2$$

Similarly, we find

$$(39b) \quad E \left[\frac{d^2 F}{\delta g_r^2} \Big|_{g=\gamma} \right] = E \left[\frac{2}{k} \left(\sum_i^k c_{ir} x_i^2 + \sum_i^k c_{ir} K_i (-y_{ir}^2 + \right. \right. \\ \left. \left. + 4 \gamma_r y_{ir} x_i - 4 \gamma_r^2 x_i^2) \right) \right] \\ = \frac{2}{k} E \left[\sum_i^k c_{ir} (\mu_i + \bar{\epsilon}_i)^2 + \right. \\ \left. + \sum_i^k c_{ir} K_i (-(\gamma_r \mu_i + \epsilon_{ir})^2 + 4 \gamma_r x_i \delta_{ir}) \right]$$

$$= \frac{2}{k} \sum_i^k c_{ir} [\bar{\mu}_i^2 + K_i \sigma^2 - K_i (\gamma_r^2 \bar{\mu}_i^2 + \sigma^2)]$$

$$= \frac{2}{k} \sum_i^k c_{ir} \bar{\mu}_i^2 (1 - \gamma_r^2 K_i)$$

Comment

From the assumption of stationarity for the variances and covariances of the first order derivatives, we conclude that the expectations of the second order derivatives are also stationary.

ii) Conditions for the second order derivatives

In order to obtain unique estimates from complete data, it is necessary to assume that the two largest eigenvalues of the covariance matrix of y_{ij} are not equal. Thus, in this case, the function F defined by (23) has a minimum for $g = \gamma$ when k tends to infinity. If the data are incomplete, we shall also use the assumption that F has a minimum at $g = \gamma$. But, in order to obtain the large sample properties, we need the additional condition that a Taylor expansion using the first and second order derivatives of F is sufficient for studying ΔF at the point $g = \gamma$. The matrix formed by the second order partial derivatives of F would be positive definite at $g = \gamma$ were it not for the fact that the normalizing conditions which has been imposed on the loadings makes it singular. We must therefore reformulate the conditions for uniqueness in the following manner:

The differential ΔF at the point $g = \gamma$ can be written

$$(40) \quad \Delta F = \frac{d^2 F}{2} = \frac{1}{2} \Delta' A \Delta \quad \text{where}$$

$$(40a) \quad \Delta = g - \gamma = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ . \\ . \\ \Delta_p \end{pmatrix} \quad \text{and}$$

$$(40b) \quad A = \left(\frac{d^2 F}{\delta g_r \delta g_s} \bigg|_{g = g^\star} \right) \quad \text{is the matrix of second order derivatives taken at the point } g^\star, \text{ between } g \text{ and } \gamma.$$

The normalizing condition on the loadings gives

$$(41a) \quad \gamma_p = \frac{+}{(-)} \sqrt{1 - \sum_j^{p-1} \gamma_j^2} \quad \text{and}$$

$$(41b) \quad \Delta_p = - \sum_j^{p-1} \frac{\gamma_j}{\gamma_p} \Delta_j \quad \text{for } \gamma_p \neq 0^1)$$

Let

$$(42a) \quad \Delta^\star = \begin{pmatrix} 1 & 0 & 0 & . & . & 0 \\ 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & 1 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \frac{\gamma_1}{\gamma_p} & \frac{\gamma_2}{\gamma_p} & . & . & \frac{\gamma_{p-1}}{\gamma_p} & 1 \end{pmatrix}$$

which has the obvious inverse:

$$(42b) \quad \Delta^{\star-1} = \begin{pmatrix} 1 & 0 & 0 & . & . & 0 \\ 0 & 1 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ \frac{-\gamma_1}{\gamma_p} & \frac{-\gamma_2}{\gamma_p} & . & . & \frac{-\gamma_{p-1}}{\gamma_p} & 1 \end{pmatrix}$$

¹ if $\gamma_p = 0$ we can reorder the observed variables, so that the loading for the p:th variable differs from zero.

We then have

$$(42c) \quad \Delta^{\star} \Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_{p-1} \\ 0 \end{pmatrix} = \Delta_0$$

and

$$(43) \quad \begin{aligned} \Delta F &= \frac{1}{2} \Delta' \Delta^{\star'} (\Delta^{\star-1})' A \Delta^{\star-1} \Delta^{\star} \Delta \\ &= \frac{1}{2} \Delta_0' (\Delta^{\star-1})' A \Delta^{\star-1} \Delta_0 = \frac{1}{2} \Delta_0' A^{\star} \Delta_0 \end{aligned}$$

where

$$(43a) \quad A^{\star} = (\Delta^{\star-1})' A \Delta^{\star-1}$$

Here the differential ΔF is a function of the $(p-1)$ independent variables $\Delta_1 \dots \Delta_{p-1}$.

As the last element in Δ_0 is zero, we see that the last column and the last row of $A^{\star}$ do not affect ΔF and can thus be eliminated. Let $A_{11}^{\star}$ denote the upper left part of $A^{\star}$ and Δ_{01} the vector consisting of the first $(p-1)$ elements of Δ_0 .

Then the differential can be rewritten as

$$(43b) \quad \Delta F = \frac{1}{2} \Delta_{01}' A_{11}^{\star} \Delta_{01}$$

and we obtain the following condition for uniqueness: There exists an $\epsilon_0 > 0$ so that

$$(44) \quad \text{plim}_{k \rightarrow \infty} (x' A_{11}^{\star} x \mid g^{\star} = \gamma) \geq (x' x) \cdot \epsilon_0 > 0$$

for every x not equal to the zero vector. In other words, $A_{11}^{\star}$ converges stochastically to a positive definite matrix at the point $g^{\star} = \gamma$.

4.3 Consistency

The previous condition implies that for every $\epsilon > 0$ there exists an integer k_0 and an $\epsilon_1 > 0$ such that, for every $k > k_0$ and every x not equal to the zero vector, we have

$$(45a) \quad P(x' A_{11}^{\star} x > (x' x) \cdot \epsilon_1 \mid g^{\star} = \gamma) \geq 1 - \epsilon$$

Furthermore, as $A_{11}^{\star}$ is a continuous function of $g^{\star}$, it follows that there exists a region R_{γ} , with the point γ as an inner point, such that

$$(45b) \quad P(x' A_{11}^{\star} x > (x' x) \cdot \epsilon_1 \mid g^{\star} \in R_{\gamma}) \geq 1 - 2\epsilon$$

Because $\left[\frac{dF}{\delta g_r} \right]_{g=\gamma}$ converges to zero when $k \rightarrow \infty$, there exists for every $\delta > 0$ an integer k_1 such that

$$(45c) \quad P\left(\left| \frac{dF}{\delta g_r} \right| < \delta \cdot \epsilon_1 \mid g = \gamma \right) \geq 1 - \epsilon \quad \text{for every } k > k_1$$

Provide that δ is chosen so that every g , for which $|g - \gamma| < \delta$, belongs to R_{γ} , it follows that for $k > \max(k_0, k_1)$

$$(45d) \quad P\left(\frac{dF}{\delta g_r} = 0 \mid \max |g - \gamma| < \delta \right) \geq 1 - 3\epsilon$$

As δ and ϵ can be chosen arbitrarily small, we conclude that the MO estimates are consistent.

4.4 Large sample variances and covariances of the loadings in terms of theoretical parameters

Using a Taylor expansion of $\frac{dF}{\delta g}$ ¹⁾ around $g = \gamma$ gives

$$(46) \quad \frac{dF}{\delta g} = \left. \frac{dF}{\delta g} \right|_{g=\gamma} + \left(\left. \frac{d^2 F}{\delta g_r \delta g_s} \right|_{g=g^{\star}} \right) \cdot (g - \gamma)$$

¹⁾ $\frac{dF}{\delta g}$ denotes the column vector with the elements $\frac{dF}{\delta g_r}$.

where $g^\star$ lies between g and γ .

Using (33b), (40a) and (40b), this can thus be written as

$$(46a) \quad \frac{dF}{\delta g} = y + A \Delta$$

It then follows from (42c) and (43a) that we can rewrite (46) as

$$(46b) \quad (\Delta^{\star-1})' \frac{dF}{\delta g} = (\Delta^{\star-1})' y + (\Delta^{\star-1})' A \Delta^{\star-1} \Delta^\star \Delta$$

$$= (\Delta^{\star-1})' y + A^\star \Delta_0 = y_0 + A^\star \Delta_0$$

where

$$(46c) \quad y_0 = (\Delta^{\star-1})' y$$

As $\frac{dF}{\delta g}$ is always zero in the solution and as the last element of Δ_0 is zero we can eliminate the last column and the last row of $A^\star$.

Thus, letting ϕ denote the zero vector and $A_{11}^\star$ the upper left part of $A^\star$, we have

$$(47a) \quad \phi = y_{01} + A_{11}^\star \Delta_{01} \quad \text{where } y_{01} \text{ is a vector consisting of the first } (p-1) \text{ elements of } y_0$$

or

$$(47b) \quad \Delta_{01} = -A_{11}^{\star-1} y_{01} \quad \text{for } A_{11}^\star \text{ nonsingular}$$

or

$$(47c) \quad \sqrt{k} \Delta_{01} = -A_{11}^{\star-1} (\sqrt{k} y_{01})$$

But $\sqrt{k} y$ is asymptotically normally distributed with covariance matrix Φ (see 34 a-b). Thus, the asymptotic distribution of

$$(48) \quad (\Delta^{\star-1})' \sqrt{k} y$$

will also be normal with covariance matrix

$$(48a) \quad (\Delta^{\star-1})' \Phi \Delta^{\star-1} = \Phi^{\star}$$

If we eliminate the last row and column of $\Phi^{\star}$, we obtain the covariance matrix of $\sqrt{k} y_{01}$, denoted by $\Phi_{11}^{\star}$. Thus $\sqrt{k} y_{01}$ is asymptotically normally distributed with covariance matrix $\Phi_{11}^{\star}$. As the estimates are consistent, g will tend to γ as $k \rightarrow \infty$. Because it is situated between g and γ , $g^{\star}$ will also have the same limit γ . Thus, $A_{11}^{\star}$ will tend to a constant non-singular matrix. According to Cramér (1945), $\sqrt{k} \Delta_{01}$ will then be asymptotically normally distributed with covariance matrix

$$(49) \quad A_{11}^{\star-1} \Phi_{11}^{\star} (A_{11}^{\star-1})'$$

Define

$$(50) \quad \Delta^{\star\star} = \begin{pmatrix} 1 & 0 & 0 & \cdot & 0 \\ 0 & 1 & 0 & \cdot & 0 \\ 0 & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -\gamma_1 & -\gamma_2 & \cdot & \cdot & -\gamma_{p-1} \\ \gamma_p & \gamma_p & \cdot & \cdot & \gamma_p \end{pmatrix} \quad \left. \begin{array}{l} \Delta^{\star\star} \text{ is of order} \\ p \times (p-1) \end{array} \right\}$$

then

$$(50a) \quad \Delta = \Delta^{\star\star} \Delta_{01}$$

and the asymptotic distribution of Δ is normal with covariance matrix

$$(51) \quad \psi = \frac{1}{k} \Delta^{\star\star} A_{11}^{\star-1} \Phi_{11}^{\star} (\Delta^{\star\star} A_{11}^{\star-1})'$$

In conclusion, the asymptotic distribution of the loadings is normal with covariance matrix ψ .

Comments

$A_{11}^{\star}$ is the upper left part of $A^{\star}$ which is given by (43a). According to (44), $A_{11}^{\star}$ is nonsingular except for zero probabilities. $A^{\star}$ is given by (43a) as

$$(52) \quad A^{\star} = (\Delta^{\star-1})' A \Delta^{\star-1}$$

Because the method is consistent, the elements of A are given by (39a, b) and (40b). As the loadings are normalized, it follows immediately that the second order moments for the loadings exist for finite samples.

When the data are complete, the covariance matrix for the loadings has been given by Anderson (1963), Lyttkens (see Wold, 1966b) and others. In this case, the variances and covariances can be written in a much simpler form as compared to the case with missing observations.

The variances for the loadings are

$$(53) \quad D^2(g_j) = \frac{\sigma^2 (\sigma_{\mu}^2 + \sigma^2) (1 - \gamma_1^2)}{k \sigma_{\mu}^4}$$

If there are no missing observations, formula (51) gives the same variances as (53). To illustrate this, let us consider the case with only two variables, i.e. $p = 2$. We then find

$$a) \quad \Phi = 4(\sigma_{\mu}^2 + \sigma^2) \begin{pmatrix} 1 - \gamma_1^2 & -\gamma_1 \gamma_2 \\ -\gamma_1 \gamma_2 & 1 - \gamma_2^2 \end{pmatrix} \cdot \sigma^2$$

$$b) \quad A = 2 \sigma_{\mu}^2 \cdot \begin{pmatrix} 1 - \gamma_1^2 & -\gamma_1 \gamma_2 \\ -\gamma_1 \gamma_2 & 1 - \gamma_2^2 \end{pmatrix}$$

$$c) \quad \Delta^{\star} = \begin{pmatrix} 1 & 0 \\ \frac{\gamma_1}{\gamma_2} & 1 \end{pmatrix}$$

$$d) \quad \Delta^{\star^{-1}} = \begin{pmatrix} 1 & 0 \\ -\frac{\gamma_1}{\gamma_2} & 1 \end{pmatrix}$$

This gives

$$(54) \quad A^{\star} = (\Delta^{\star^{-1}})' A \Delta^{\star^{-1}}$$

$$= 2\sigma_{\mu}^2 \begin{pmatrix} 1 + \gamma_1^2 + \frac{\gamma_1^4}{\gamma_2^2} & -\gamma_1 \gamma_2 - \frac{\gamma_1^3}{\gamma_2^2} \\ -\gamma_1 \gamma_2 - \frac{\gamma_1^3}{\gamma_2^2} & 1 - \gamma_2^2 \end{pmatrix}$$

and

$$(54a) \quad A_{11}^{\star} = 2 \sigma_{\mu}^2 (1 + \gamma_1^2 + \frac{\gamma_1^4}{\gamma_2^2})$$

In the same way we find that

$$(55) \quad \Phi_{11}^{\star} = 4(\sigma_{\mu}^2 + \sigma^2) \sigma^2 (1 + \gamma_1^2 + \frac{\gamma_1^4}{\gamma_2^2})$$

and that

$$(56) \quad A_{11}^{\star^{-1}} \Phi_{11}^{\star} (A_{11}^{\star^{-1}})' = \frac{(\sigma_{\mu}^2 + \sigma^2) \sigma^2}{\sigma_{\mu}^4 (1 + \gamma_1^2 + \frac{\gamma_1^4}{\gamma_2^2})}$$

From the normalizing condition it follows that

$$(56b) \quad 1 + \gamma_1^2 + \frac{\gamma_1^4}{\gamma_2^2} = \frac{(1 + \gamma_1^2)(1 - \gamma_1^2) + \gamma_1^4}{1 - \gamma_1^2} = \frac{1}{1 - \gamma_1^2}$$

Thus

$$(56c) \quad A_{11}^{\star-1} \Phi_{11}^{\star} (A_{11}^{\star-1})' = \frac{(\sigma_{\mu}^2 + \sigma^2) \sigma^2 (1 - \gamma_1^2)}{\sigma_{\mu}^4}$$

In addition,

$$(57) \quad \Delta^{\star\star} = \begin{pmatrix} 1 \\ -\gamma_1 \\ \gamma_2 \end{pmatrix}$$

and finally, we obtain

$$(58) \quad \psi = \frac{1}{k} \Delta^{\star\star} A_{11}^{\star-1} \Phi_{11}^{\star} (A_{11}^{\star-1})' \Delta^{\star\star'} \\ = \frac{(\sigma_{\mu}^2 + \sigma^2) \sigma^2}{k \sigma_{\mu}^4} \begin{pmatrix} 1 - \gamma_1^2 & -\gamma_1 \gamma_2 \\ -\gamma_1 \gamma_2 & 1 - \gamma_2^2 \end{pmatrix}$$

from which we derive the variance for the loadings:

$$(59) \quad D^2(g_j) = \frac{(\sigma_{\mu}^2 + \sigma^2) \sigma^2 (1 - \gamma_j^2)}{k \sigma_{\mu}^4}$$

4.5 Convergence of the iterative procedure

It is not possible to show that the procedure (17a-c) always converges to a non degenerate solution, given that we have begun with an arbitrary first

approximation. This point was mentioned earlier in section three. However, from condition (44) and the consistency of the method it follows, with probability one, that the sum of squared residuals has a minimum in the neighbourhood of $g = \gamma$. Thus, the iterative procedure will converge to a unique solution of the system (15a-c) for an arbitrary first approximation of the loadings in the neighbourhood of the point $g = \gamma$.

4.6 Limiting distribution of the estimated component and residuals

In this section we shall show that the limiting distributions of the estimated component and residuals are those of the corresponding theoretical quantities. We shall also consider a two-part decomposition of the sum of squared observations and the limiting distributions of these parts.

i) Asymptotic distribution of the estimated component

In (15b), the estimated component is defined as

$$(60) \quad x_i = \frac{\sum_j^p c_{ij} g_j y_{ij}}{\sum_j^p c_{ij} g_j^2}$$

As the MO estimates are consistent, the loadings g_j tend to the corresponding theoretical loadings γ_j . According to Cramér (1945, theorem 20.6) the distribution of x_i tends to the distribution of

$$(61a) \quad \frac{\sum_j^p c_{ij} \gamma_j y_{ij}}{\sum_j^p c_{ij} \gamma_j^2}$$

By (12) the theoretical component is defined as

$$(61b) \quad \xi_i = \frac{\sum_j^p c_{ij} \gamma_j y_{ij}}{\sum_j^p c_{ij} \gamma_j^2}$$

Thus, the limiting distribution of x_i is the same as that of ξ_i .

ii) Asymptotic distribution of the estimated residuals

The estimated residuals are

$$(62) \quad e_{ij} = (y_{ij} - g_j x_i)$$

Using the same theorem as above, we find that the limiting distribution of e_{ij} is that of $(y_{ij} - \gamma_j \xi_i)$. By (13),

$$(62a) \quad \delta_{ij} = y_{ij} - \gamma_j \xi_i$$

and we conclude that the asymptotic distribution of e_{ij} is the same as that of δ_{ij} .

iii) Decomposition of the sum of squared observations

From the specification of the model (10 - 13) it is easily seen that

$$\begin{aligned} (63a) \quad & \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} y_{ij}^2 \\ &= \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} \gamma_j^2 \xi_i^2 + \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} \delta_{ij}^2 \end{aligned}$$

The corresponding decomposition in terms of the estimated quantities is

$$\begin{aligned} (63b) \quad & \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} y_{ij}^2 \\ &= \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} g_j^2 x_i^2 + \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} e_{ij}^2 \end{aligned}$$

which follows directly from (14 - 14b).

To show that the limiting distributions of the sums on the right side of (63b) are those of the corresponding sums in (63a) we start by considering

$$(64) \quad \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} g_j^2 x_i^2$$

We have

$$(64a) \quad \sum_j^p c_{ij} g_j^2 x_i^2 = \frac{(\sum_j^p c_{ij} g_j y_{ij})^2}{\sum_j^p c_{ij} g_j^2}$$

$$= \frac{\sum_j^p c_{ij} g_j^2 y_{ij}^2 + 2 \sum_{j < s} \sum_s c_{ij} c_{is} g_j g_s y_{ij} y_{is}}{\sum_j^p c_{ij} g_j^2}$$

From the specification of the model (10 – 13) we find

$$(64b) \quad E(y_{ij}^2) = \gamma_j^2 \bar{\mu}_i^2 + \sigma^2$$

$$(64c) \quad E(y_{ij} y_{is}) = \gamma_j \gamma_s \bar{\mu}_i^2$$

where $E(\mu_i^2) = \bar{\mu}_i^2$

Using this we rewrite (64a) as

$$(65a) \quad \sum_j^p c_{ij} g_j^2 x_i^2 = \frac{1}{\sum_j^p c_{ij} g_j^2} [\sum_j^p c_{ij} g_j^2 (y_{ij}^2 - \gamma_j^2 \bar{\mu}_i^2 - \sigma^2) +$$

$$\begin{aligned}
& + 2 \sum_{j < s} \sum c_{ij} c_{is} g_j g_s (y_{ij} y_{is} - \gamma_j \gamma_s \bar{\mu}_i^2)] + \\
& + \frac{1}{\sum_j^p c_{ij} g_j^2} [\sum_j^p c_{ij} g_j^2 (\gamma_j^2 \bar{\mu}_i^2 + \sigma^2) + \\
& + 2 \sum_{j < s} \sum c_{ij} c_{is} g_j g_s \gamma_j \gamma_s \bar{\mu}_i^2] = A_i^g + B_i^g
\end{aligned}$$

where B_i^g can be rewritten as

$$(65b) \quad B_i^g = \frac{\bar{\mu}_i^2}{\sum_j^p c_{ij} g_j^2} (\sum_j^p c_{ij} g_j \gamma_j)^2 + \sigma^2$$

Similarly, for $\sum c_{ij} \gamma_j^2 \xi_i^2$, we obtain

$$\begin{aligned}
(65c) \quad \sum_j^p c_{ij} \gamma_j^2 \xi_i^2 &= \frac{1}{\sum_j^p c_{ij} \gamma_j^2} [\sum_j^p c_{ij} \gamma_j^2 (y_{ij}^2 - \gamma_j^2 \bar{\mu}_i^2 - \sigma^2) + \\
& + 2 \sum_{j < s} \sum \gamma_j \gamma_s (y_{ij} y_{is} - \gamma_j \gamma_s \bar{\mu}_i^2)] + \\
& + \frac{\bar{\mu}_i^2}{\sum_j^p c_{ij} \gamma_j^2} (\sum_j^p c_{ij} \gamma_j^2)^2 + \sigma^2 = A_i^\gamma + B_i^\gamma
\end{aligned}$$

Then

$$(66a) \quad \sum_j^p c_{ij} g_j^2 x_i^2 - \sum_j^p c_{ij} \gamma_j^2 \xi_i^2 = (A_i^g - A_i^\gamma) + (B_i^g - B_i^\gamma)$$

where

$$(66b) \quad A_i^g - A_i^\gamma = \sum_j^p c_{ij} (y_{ij}^2 - \gamma_j^2 \bar{\mu}_i^2 - \sigma^2) \left(\frac{g_j^2}{\sum_r^p c_{ir} g_r^2} - \frac{\gamma_j^2}{\sum_r^p c_{ir} \gamma_r^2} \right) + 2 \sum_{j < s} c_{ij} c_{is} (y_{ij} y_{is} - \gamma_j \gamma_s \bar{\mu}_i^2) \left(\frac{g_j g_s}{\sum_r^p c_{ir} g_r^2} - \frac{\gamma_j \gamma_s}{\sum_r^p c_{ir} \gamma_r^2} \right)$$

and

$$(66c) \quad B_i^g - B_i^\gamma = \bar{\mu}_i^2 \left(\frac{\left(\sum_j^p c_{ij} g_j \gamma_j \right)^2}{\sum_j^p c_{ij} g_j^2} - \sum_j^p c_{ij} \gamma_j^2 \right)$$

Consider now all possible realisations of the set $c_{i1}, c_{i2}, \dots, c_{ip}$. Because p is a given fixed integer, the total number of realisations, k_p is limited. Let n_s denote the number of times the s th realisation occurs in (64). From condition (35a,b) it follows that, for every s , $\lim_{k \rightarrow \infty} \frac{n_s}{k}$ will exist, and that this limiting value will be contained in the closed interval $[0, 1]$.

After a suitable rearrangement of the terms in (64), we find

$$(67) \quad \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} g_j^2 x_i^2 - \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} \gamma_j^2 \xi_i^2 \\ = \frac{1}{\sqrt{k}} \sum_i^k [(A_i^g - A_i^\gamma) + (B_i^g - B_i^\gamma)]$$

$$\begin{aligned}
&= \frac{1}{\sqrt{k}} \sum_{i=1}^{N_1} [(A_i^g - A_i^\gamma) + (B_i^\gamma - B_i^\gamma)] + \\
&+ \sum_{i=N_1+1}^{N_2} [(A_i^g - A_i^\gamma) + (B_i^\gamma - B_i^\gamma)] + \\
&\dots + \sum_{i=N_{k_p-1}+1}^{N_{k_p}} [(A_i^g - A_i^\gamma) + (B_i^g - B_i^\gamma)]
\end{aligned}$$

$$\text{where } N_s = \sum_{i=1}^s n_i$$

In (67) each partial sum corresponds to the same realisation of $c_{i1}, c_{i2}, \dots, c_{ip}$. For each of these partial sums we have:

$$\begin{aligned}
(68a) \quad & \frac{1}{\sqrt{k}} \sum_{i=1}^{N_{s+1}} (A_i^g - A_i^\gamma) \\
&= \frac{1}{\sqrt{k}} \left[\sum_j^p c_{ij} \left(\frac{g_j^2}{\sum_r^p c_{ir} g_r^2} - \frac{\gamma_j^2}{\sum_r^p c_{ir} \gamma_r^2} \right) \sum_{i=N_s+1}^{N_{s+1}} c_{ij} (y_{ij}^2 - \gamma_j^2 \bar{\mu}_i^2 - \sigma^2) \right] \\
&+ \frac{2}{\sqrt{k}} \left[\sum_{j < t}^p \sum_{i=t}^p c_{ij} c_{it} \left(\frac{g_j g_t}{\sum_r^p c_{ir} g_r^2} - \frac{\gamma_j \gamma_t}{\sum_r^p c_{ir} \gamma_r^2} \right) \right] \\
&\sum_{i=N_s+1}^{N_{s+1}} c_{ij} c_{it} (y_{ij} y_{it} - \gamma_j \gamma_t \bar{\mu}_i^2)]
\end{aligned}$$

Now,

$$(68b) \quad \frac{1}{k} \sum_{i=N_s+1}^{N_s+1} c_{ij} (y_{ij}^2 - \gamma_j^2 \bar{\mu}_i^2 - \sigma^2)$$

and

$$(68c) \quad \frac{1}{k} \sum_{i=N_s+1}^{N_s+1} c_{ij} c_{it} (y_{ij} y_{it} - \gamma_j \gamma_t \bar{\mu}_i^2)$$

will tend to zero as k tends to infinity. Because the asymptotic distribution of g_j has expectation equal to γ_j and variance of order $\frac{1}{k}$ it follows that

$$(68d) \quad \sqrt{k} \left(\frac{g_j^2}{\sum_r^p c_{ir} g_r^2} - \frac{\gamma_j^2}{\sum_r^p c_{ir} \gamma_r^2} \right)$$

and

$$(68e) \quad \sqrt{k} \left(\frac{g_j g_t}{\sum_r^p c_{ir} g_r^2} - \frac{\gamma_j \gamma_t}{\sum_r^p c_{ir} \gamma_r^2} \right)$$

will tend to distributions with zero expectations and finite variances as k tends to infinity. We conclude that

$$(68f) \quad \frac{1}{\sqrt{k}} \sum_i^k (A_i^g - A_i^\gamma)$$

tends to zero as k tends to infinity.

Turning to the partial sums of $(B_i^g - B_i^\gamma)$ we find from (66c) that

$$(69a) \quad \frac{1}{\sqrt{k}} \sum_{N_s+1}^{N_s+1} (B_i^g - B_i^\gamma) =$$

$$\begin{aligned}
&= \frac{1}{\sqrt{k}} \sum_{i=N_s+1}^{N_{s+1}} \bar{\mu}_i^2 \left(\frac{\left(\sum_j^p c_{ij} g_j \gamma_j \right)^2}{\sum_j^p c_{ij} g_j^2} - \sum_j^p c_{ij} \gamma_j^2 \right) \\
&= \sqrt{k} \left(\frac{\left(\sum_j^p c_{ij} g_j \gamma_j \right)^2}{\sum_j^p c_{ij} g_j^2} - \sum_j^p c_{ij} \gamma_j^2 \right) \frac{1}{k} \sum_{i=N_s+1}^{N_{s+1}} \bar{\mu}_i^2
\end{aligned}$$

The second part of (69a), i.e.,

$$(69b) \quad \frac{1}{k} \sum_{i=N_s+1}^{N_{s+1}} \bar{\mu}_i^2$$

is bounded above as the number of terms in the sum is less than or equal to k . On the other hand it can be shown that

$$(69c) \quad C = \sqrt{k} \left(\frac{\left(\sum_j^p c_{ij} g_j \gamma_j \right)^2}{\sum_j^p c_{ij} g_j^2} - \sum_j^p c_{ij} \gamma_j^2 \right) = \sqrt{k} C_0$$

will tend to a distribution with zero mean and zero variance when k tends to infinity.

Following Cramér (1945) we find the asymptotic variance of C :

$$\begin{aligned}
(69d) \quad D^2(C) &= k \left[\sum_j^p c_{ij} \left[\frac{\partial C_0}{\partial g_j} \right]_{g=\gamma} \right]^2 D^2(g_j) + \\
&\quad + 2 \sum_{j < s} \sum c_{ij} c_{is} \left[\frac{\partial C_0}{\partial g_j} \right]_{g=\gamma} \left[\frac{\partial C_0}{\partial g_s} \right]_{g=\gamma} \text{cov}(g_j g_s)]
\end{aligned}$$

But as $\frac{\partial C_0}{\partial g_j}$ is zero at the point $g = \gamma$, it follows that the asymptotic variance of C_0 is zero to the order of $\frac{1}{k}$. Thus C converges in probability to zero and the same is true for

$$(69e) \quad \frac{1}{\sqrt{k}} \sum_i^k (B_i^g - B_i^\gamma)$$

In conclusion, we have proved that the limiting distribution of

$$(70a) \quad \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} g_j^2 x_i^2$$

is that of

$$(70b) \quad \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} \gamma_j^2 \xi_i^2$$

It follows immediately that the limiting distribution of

$$(71a) \quad \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} e_{ij}^2$$

is that of

$$(71b) \quad \frac{1}{\sqrt{k}} \sum_i^k \sum_j^p c_{ij} \delta_{ij}^2$$

4.7 Estimation of large sample variances and covariances of the loadings in terms of estimated parameters

The variances and covariances can be estimated from the sample. Using the results obtained in the preceding section, we proceed as follows:

The residual for the component representation of the model can be written:

$$(72) \quad \delta_{ij} = \epsilon_{ij} - \gamma_j \frac{\sum_s^p c_{is} \gamma_s \epsilon_{is}}{\sum_s^p c_{is} \gamma_s^2} = \epsilon_{ij} - \gamma_j \sum_s^p c_{is} \beta_s \epsilon_{is}$$

where

$$(72a) \quad \beta_s = \frac{\gamma_s}{\sum_r^p c_{ir} \gamma_r^2}$$

and

$$(72b) \quad \sum_s^p c_{is} \beta_s^2 = \frac{1}{\sum_r^p c_{ir} \gamma_r^2}$$

We then find that

$$\begin{aligned} (73) \quad \sum_j^p c_{ij} \delta_{ij}^2 &= \sum_j^p c_{ij} (\epsilon_{ij} - \gamma_j \sum_s^p c_{is} \beta_s \epsilon_{is})^2 \\ &= \sum_j^p c_{ij} \epsilon_{ij}^2 - 2 \sum_j^p c_{ij} \gamma_j \epsilon_{ij} \sum_j^p c_{is} \beta_s \epsilon_{is} + \\ &\quad + \sum_j^p c_{ij} \gamma_j^2 \left(\sum_s^p c_{is} \beta_s \epsilon_{is} \right)^2 \\ &= \sum_j^p c_{ij} \epsilon_{ij}^2 - 2 \left(\sum_s^p c_{is} \beta_s \epsilon_{is} \right)^2 \cdot \sum_j^p c_{ij} \gamma_j^2 + \\ &\quad + \sum_j^p c_{ij} \gamma_j^2 \left(\sum_s^p c_{is} \beta_s \epsilon_{is} \right)^2 \\ &= \sum_j^p c_{ij} \epsilon_{ij}^2 - \sum_j^p c_{ij} \gamma_j^2 \left(\sum_s^p c_{is} \beta_s \epsilon_{is} \right)^2 \end{aligned}$$

or

$$(73a) \quad \sum_j^p c_{ij} \epsilon_j^2 = \sum_j^p c_{ij} \delta_{ij}^2 + \sum_j^p c_{ij} \gamma_j^2 \left(\sum_s^p c_{is} \beta_s \epsilon_{is} \right)^2$$

The corresponding ranks for this equation are

$$(74) \quad \sum_j^p c_{ij} = (\sum_j^p c_{ij} - 1) + 1$$

If the residuals in the factor representation of the model are normally distributed, it immediately follows that

$$(75) \quad \frac{\sum_j^p c_{ij} \epsilon_{ij}^2}{\sigma^2}$$

is chi-square distributed with $\sum_j^p c_{ij}$ degrees of freedom and that

$$(76) \quad \frac{\sum_i^k \sum_j^p c_{ij} \delta_{ij}^2}{\sigma^2}$$

is distributed as chi-square with

$$(76a) \quad f = (\sum_i^k \sum_j^p c_{ij} - k)$$

degrees of freedom. From section 4.6, we conclude that

$$(77) \quad \frac{\sum_i^k \sum_j^p c_{ij} e_{ij}^2}{\sigma^2}$$

is asymptotically distributed as chi-square with f degrees of freedom.

The estimate of the residual variance is given by

$$(78) \quad \hat{\sigma}^2 = \frac{\sum_i^k \sum_j^p c_{ij} e_{ij}^2}{f}$$

It is common practice to reduce the number of degrees of freedom by the effective number of estimated parameters, which in this case is $(p - 1)$.

We may then use

$$(79) \quad f^{\star} = f - (p - 1)$$

instead of f in (78) and estimate the residual variance with

$$(79a) \quad \hat{\sigma}^2 = \frac{\sum_i^k \sum_j^p c_{ij} e_{ij}^2}{f^{\star}}$$

To estimate the covariance matrix for the loadings we shall also need estimates of the covariance matrix for the first order derivatives (33b). From the specification of the model it follows that

$$(80) \quad E(\xi_i^2) = \bar{\mu}_i^2 + K_i \sigma^2$$

By section 4.6 we find that asymptotically,

$$(80a) \quad E(x_i^2) = \bar{\mu}_i^2 + K_i \sigma^2$$

Using this, we estimate the variance of the first order derivatives as

$$(81) \quad \frac{1}{k} \hat{\varphi}_{rr} = \frac{4 \hat{\sigma}^2}{k} \left(\frac{1}{k} \sum_i^k c_{ir} (1 - \hat{K}_i g_r^2) x_i^2 \right)$$

where

$$(81a) \quad \hat{K}_i = \frac{1}{\sum_j^p c_{ij} g_j^2}$$

and the covariance of the derivatives as

$$(82) \quad \frac{1}{k} \hat{\varphi}_{rs} = \frac{-4 \hat{\sigma}^2}{k} \left(\frac{1}{k} \sum_i^k c_{ir} c_{is} g_r g_s \cdot \hat{K}_i \cdot x_i^2 \right)$$

The second order derivatives are estimated by

$$(83a) \quad \hat{a}_{rr} = \frac{2}{k} \sum_i^k c_{ir} (x_i^2 (1 - g_r^2 \hat{K}_i) - \hat{\sigma}^2 (\hat{K}_i - g_r^2 \hat{K}_i^2))$$

and

$$(83b) \quad \hat{a}_{rs} = \frac{-2 g_r g_s}{k} \sum_i^k c_{ir} c_{is} (\hat{K}_i x_i^2 - \hat{K}_i^2 \hat{\sigma}^2)$$

The operators $\Delta^\star$ and $\Delta^{\star\star}$ are estimated in the usual way, by substituting the estimated loadings for the corresponding true loadings.

Using the above, we obtain the estimated covariance matrix for the loadings in a way analogous to that used for (51).

$$(84) \quad \hat{\psi} = \frac{1}{k} \hat{\Delta}^{\star\star} \hat{A}_{11}^{\star-1} \hat{\Phi}_{11}^\star (\hat{\Delta}^{\star\star} \hat{A}_{11}^{\star-1})'$$

where $\wedge$ denotes an estimate of the corresponding theoretical quantity. Using the results obtained in section 4.6, it follows immediately that

$$(85) \quad \text{plim}_{k \rightarrow \infty} k \hat{\psi} = \lim_{k \rightarrow \infty} k \psi$$

4.8 Confidence regions for the loadings

We may use (84) to obtain approximate confidence regions for the true loadings. As the loadings tend to a normal distribution, it follows that (asymptotically)

$$(86) \quad P \{ (g_0 - \gamma_0)' \psi_0^{-1} (g_0 - \gamma_0) \leq \chi^2_{1-\alpha} (p-1) \} = 1 - \alpha$$

where

$$(86a) \quad g_0 = \begin{pmatrix} g_1 \\ \vdots \\ g_{p-1} \end{pmatrix} \quad \gamma_0 = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_{p-1} \end{pmatrix}$$

ψ_0 is the upper left part of the covariance matrix for the loadings, in which the last row and last column are eliminated; and $\chi^2_{1-\alpha} (p-1)$ denotes the upper $(1 - \alpha)$ critical value of the chi-square distribution with $(p - 1)$ degrees of freedom. Following Scheffé (1959), we can write, for every constant non zero vector h of order $(p - 1)$,

$$(87) \quad P \{ |h'g_0 - h'\gamma_0| \leq \sqrt{\chi^2_{1-\alpha} (p-1)} (h'\psi_0 h)^{1/2} \} \geq 1 - \alpha$$

By choosing $h' = (1, 0, 0 \dots 0)$ we find that

$$(87a) \quad P \{ |g_1 - \gamma_1| \leq \sqrt{\chi^2_{1-\alpha} (p-1) \cdot \psi_{11}} \} \geq 1 - \alpha$$

where ψ_{11} is the variance for g_1 .

In general, we find

$$(87b) \quad P \{ |g_j - \gamma_j| \leq \sqrt{\chi^2_{1-\alpha} (p-1) \cdot \psi_{jj}} \} \geq 1 - \alpha$$

The confidence interval defined by (87b) holds simultaneously for all j with a confidence level of at least $1 - \alpha$.

If the matrix ψ is not known we can use the estimated covariance matrix, i.e. we use $\hat{\psi}_{jj}$ instead of ψ_{jj} in (87b). This approximation is justified by (85) which implies that the estimated covariance matrix tends to the "true" covariance matrix ψ as k goes to infinity.

4.9 Testing of the model

For complete data the problem of testing the model, i.e. testing the hypothesis that there is one factor (component) against the alternative that there are more than one, has been treated by Bartlett (1950), Anderson (1963), and others. The techniques used are based on the eigenvalues of the covariance matrix for the observations. For the case of incomplete data, the situation is different as the covariance matrix cannot be calculated. Therefore, an appropriate test seems to be extremely complicated to derive. One possibility would be to apply some nonparametric test on the residuals in the sample, for example a run test, as used in regression analysis with only one observation on the dependent variable for each value of the independent variable. However, nonparametric tests usually have rather low power.

The situation is not as limited as it appears at first sight. In many cases the model can be derived from the theories of the field of application. The interest will then focus on the parameters of the model (Wold, 1956). In psychology, for example, the model and its unknown parameters are expressed in terms of psychological theories. From these theories and from experience we often know the approximate size of the reliabilities for the different variables. A common case is that these reliabilities are between 0.7 and 0.9, implying that the ratio of explained to total variation is about 0.5 to 0.8. In the next section we shall derive an estimate of this ratio which then can be used to test the hypothesis about the size of these reliabilities.

5. EXPLAINED AND UNEXPLAINED VARIATION

In certain types of applications of the one factor model, interest focuses not so much upon the loadings but upon the variance explained by the factor. For complete data the relation between explained and unexplained variation has been investigated by several authors including Anderson (1963) who considered the ratio between unexplained and explained variation. In this section we derive an estimate of the ratio of explained to total variability for the case of incomplete data. This ratio is analogous to the square of the multiple correlation coefficient, the customary measure in regression analysis. We shall consider cases with normally distributed residuals and stochastic or nonstochastic factor.

From the specification of the model (2) it is easily seen that the ratio of explained to total variability is

$$(88) \quad \bar{R}^2 = \frac{\sigma_\mu^2}{\sigma_\mu^2 + p \sigma^2}$$

where

$$(88a) \quad \sigma_\mu^2 = \frac{1}{k} \sum_i^k \bar{\mu}_i^2$$

with

$$\bar{\mu}_i^2 = E(\mu_i^2)$$

$\bar{R}^2$ is a measure of the amount of variability explained by the factor in the model. By defining

$$(89) \quad \bar{R}_j^2 = \frac{\gamma_j^2 \sigma_\mu^2}{\gamma_j^2 \sigma_\mu^2 + \sigma^2}$$

we determine another measure, one related to the individual tests.

Direct estimation of these parameters leads to estimates with rather complicated properties. To avoid this we shall consider the following parameters,

which contain the same amount of information:

i) the ratio between explained and unexplained variation;

$$(90) \quad \alpha = \frac{\sigma_{\mu}^2}{p \sigma^2}$$

ii) the ratio between explained and unexplained variation for the individual tests;

$$(91) \quad \alpha_j = \frac{\gamma_j^2 \sigma_{\mu}^2}{\sigma^2} = \gamma_j^2 p \alpha$$

It is easily seen that these parameters are related in a one to one manner to $\bar{R}^2$ and $\bar{R}_j^2$ respectively

$$(92a) \quad \bar{R}^2 = \frac{\alpha}{1 + \alpha}$$

$$(92b) \quad \bar{R}_j^2 = \frac{\alpha_j}{1 + \alpha_j}$$

If $\hat{\alpha}$ is an estimate of α we obtain an estimate of $\bar{R}^2$ given by

$$(93a) \quad R^2 = \frac{\hat{\alpha}}{1 + \hat{\alpha}}$$

In the same way, for $\bar{R}_j^2$, we have

$$(93b) \quad R_j^2 = \frac{\hat{\alpha}_j}{1 + \hat{\alpha}_j}$$

where $\hat{\alpha}_j$ is given by

$$(93c) \quad \hat{\alpha}_j = g_j^2 p \hat{\alpha}$$

To obtain an estimate of α we proceed as follows:

Define the variable u according to

$$(94) \quad u = \frac{1}{k \sigma^2} \sum_i^k \xi_i^2 \sum_j^p c_{ij} \gamma_j^2 = \frac{1}{k} \sum_i^k \frac{\xi_i^2}{K_i \sigma^2}$$

where K_i is defined as in section four.

From the specification of the model (26 – 26c) we find

$$(94a) \quad \xi_i = \mu_i + K_i \sum_j^p c_{ij} \gamma_j \epsilon_{ij} = \mu_i + \bar{\epsilon}_i$$

and

$$(94b) \quad E \left(\frac{\xi_i^2}{K_i \sigma^2} \right) = \frac{\bar{\mu}_i^2}{K_i \sigma^2} + 1$$

Then

$$(95) \quad E(u) = \frac{1}{k} \sum_i^k \left(\frac{\bar{\mu}_i^2}{K_i \sigma^2} + 1 \right) = \frac{1}{\sigma^2} \sum_j^p \gamma_j^2 \frac{1}{k} \sum_i^k c_{ij} \bar{\mu}_i^2 + 1$$

Assuming that

$$(96) \quad \frac{\sum_i^k c_{ij} \bar{\mu}_i^2}{\sum_i^k c_{ij}} = \sigma_\mu^2 \quad \text{for every } j$$

and introducing f_j where

$$(96a) \quad f_j = \frac{\sum_i^k c_{ij}}{k}$$

we obtain

$$(97) \quad E(u) = \frac{\sigma_\mu^2}{\sigma^2} \sum_j^p f_j \gamma_j^2 + 1 = p \alpha \sum_j^p f_j \gamma_j^2 + 1$$

For the variance of u we shall consider two different cases:

i) Nonstochastic factor

With μ_i nonstochastic we find that

$$(98) \quad \frac{\xi_i^2}{K_i \sigma^2}$$

is distributed as noncentral chi-square with the noncentrality parameter

$$(98a) \quad \frac{\mu_i}{\sqrt{K_i} \sigma}$$

Thus,

$$(98b) \quad D^2\left(\frac{\xi_i^2}{K_i \sigma^2}\right) = 2 \left(1 + \frac{2 \bar{\mu}_i^2}{K_i \sigma^2}\right)$$

and using (96) we get

$$(98c) \quad D^2(u) = \frac{2}{k^2} \sum_i^k \left(1 + \frac{2 \bar{\mu}_i^2}{K_i \sigma^2}\right) = \frac{2}{k} \left(\frac{2 \sigma_\mu^2}{\sigma^2} \sum_j^p f_j \gamma_j^2 + 1\right) \\ = \frac{2}{k} \left(2 p \alpha \sum_j^p f_j \gamma_j^2 + 1\right)$$

ii) Stochastic factor

Assuming that the factor is normally distributed with $E(\mu_i) = 0$ and $D^2(\mu_i) = \sigma_\mu^2$, we find that

$$(99) \quad D^2(\xi_i) = \sigma_\mu^2 + K_i \sigma^2$$

and

$$\begin{aligned}
 (99a) \quad D^2 \left(-\frac{\xi_i^2}{\sigma^2 K_i} \right) &= \frac{2}{\sigma^4 K_i^2} (\sigma_\mu^2 + K_i \sigma^2)^2 \\
 &= 2 \left(\frac{\sigma_\mu^2}{K_i \sigma^2} + 1 \right)^2 = 2 \left(\frac{p \alpha}{K_i} + 1 \right)^2
 \end{aligned}$$

which gives the variance of u

$$\begin{aligned}
 (99b) \quad D^2(u) &= \frac{2}{k^2} \sum_i^k \left(\frac{p \alpha}{K_i} + 1 \right)^2 \\
 &= \frac{2}{k} \left(2 p \alpha \sum_j^p f_j \gamma_j^2 + 1 + (p \alpha)^2 \frac{1}{k} \sum_i^k \frac{1}{K_i^2} \right)
 \end{aligned}$$

From sections 4.1 and 4.7 it follows that

$$(100) \quad z = \frac{\frac{1}{k} \sum_i^k \sum_j^p c_{ij} \delta_{ij}^2}{\sigma^2}$$

is distributed independently of u and has mean and variance equal to

$$(100a) \quad E(z) = \frac{1}{k} \sum_i^k \left(\sum_j^p c_{ij} - 1 \right) = \frac{f}{k}$$

$$(100b) \quad D^2(z) = \frac{2f}{k^2}$$

Consider now the following variable:

$$(101) \quad \alpha_0 = \frac{u}{z}$$

The asymptotic mean of α_0 is

$$(101a) \quad E(\alpha_0) = \frac{E(u)}{E(z)} = \frac{p \alpha \sum_j^p f_j \gamma_j^2 + 1}{\frac{f}{k}} = \frac{kp \alpha}{f} \sum_j^p f_j \gamma_j^2 + \frac{k}{f}$$

and the asymptotic variance is

$$\begin{aligned} (101b) \quad D^2(\alpha_0) &= \frac{1}{E^2(z)} D^2(u) + \frac{E^2(u)}{E^4(z)} D^2(z) \\ &= \frac{k^2}{f^2} D^2(u) + \frac{k^4}{f^4} (p \alpha \sum_j^p f_j \gamma_j^2 + 1)^2 \frac{2f}{k^2} \\ &= \frac{k^2}{f^2} \left[D^2(u) + \frac{2(p \alpha \sum_j^p f_j \gamma_j^2 + 1)^2}{f} \right] \end{aligned}$$

where $D^2(u)$ is given by (98c) or (99b).

Using the results obtained in 4.6 we find that the variable $\hat{\alpha}_0$, given by

$$(102) \quad \hat{\alpha}_0 = \frac{\frac{1}{k} \sum_i^k x_i^2 \sum_j^p c_{ij} g_j^2}{\frac{1}{k} \sum_i^k \sum_j^p c_{ij} e_{ij}^2} = \frac{\frac{1}{k} \sum_i^k \frac{x_i^2}{\hat{K}_i}}{\frac{1}{k} \sum_i^k \sum_j^p c_{ij} e_{ij}^2}$$

with
$$\hat{K}_i = \frac{1}{\sum_j^p c_{ij} g_j^2}$$

has the same asymptotic distribution as α_0 .

In addition,

$$(103) \quad \hat{\alpha} = \frac{f \hat{\alpha}_0 - k}{p k \sum_j^p f_j g_j^2}$$

has asymptotic mean

$$(103a) \quad E(\hat{\alpha}) = \alpha = \frac{\sigma_\mu^2}{p \sigma^2}$$

and asymptotic variance

$$(103b) \quad D^2(\hat{\alpha}) = \frac{f^2}{(\sum_j^p f_j \gamma_j^2 p k)^2} D^2(\hat{\alpha}_0) + \frac{(f E(\hat{\alpha}_0) - k)^2}{(\sum_j^p f_j \gamma_j^2 p k)^4} D^2(\sum_j^p f_j g_j^2) p^2 k^2 \\ - 2\rho \frac{(f E(\hat{\alpha}_0) - k) f p k D(\hat{\alpha}_0) D(\sum_j^p f_j g_j^2)}{(\sum_j^p f_j \gamma_j^2 p k)^3}$$

where ρ denotes the correlation between α_0 and $\sum_j^p f_j g_j^2$. The exact expression for this correlation is very complicated but we get an upper limit for the variance by putting $\rho = -1$, i.e.

$$(103c) \quad D^2(\hat{\alpha}) \leq \frac{1}{(\sum_j^p f_j \gamma_j^2 p k)^2} \left[f D(\hat{\alpha}_0) + \frac{f E(\hat{\alpha}_0) - k}{\sum_j^p f_j \gamma_j^2 p k} p k D(\sum_j^p f_j g_j^2) \right]^2 \\ = \frac{1}{(\sum_j^p f_j \gamma_j^2 p k)^2} \left[f D(\hat{\alpha}_0) + p k \alpha D(\sum_j^p f_j g_j^2) \right]^2$$

where $D(\hat{\alpha}_0)$ is obtained from (101b).

The variance for $\sum_j^p f_j g_j^2$ can be obtained in the following way:
 Let f_m denote the smallest of the f_j 's.

Then

$$(104) \quad \sum_j^p f_j g_j^2 = f_m \sum_j^p g_j^2 + \sum_j^p (f_j - f_m) g_j^2 = f_m + \sum_j^p f_j^\star g_j^2$$

with $f_j^\star = f_j - f_m$

and

$$(104b) \quad D^2 \left(\sum_j^p f_j g_j^2 \right) = D^2 \left(\sum_j^p f_j^\star g_j^2 \right)$$

The asymptotic variance of $\sum_j^p f_j^\star g_j^2$ is

$$(104) \quad D^2 \left(\sum_j^p f_j^\star g_j^2 \right) = 4 \left[\sum_j^p f_j^{\star 2} D^2(g_j) + 2 \sum_{r < s} f_r^\star f_s^\star \text{cov}(g_r g_s) \right]$$

Introducing the column vector $F^\star = (f_j^\star)$, the variance of $\sum_j^p f_j g_j^2$ can be written

$$(104d) \quad D^2 \left(\sum_j^p f_j g_j^2 \right) = 4 F^{\star \prime} \psi F^\star$$

where ψ is the covariance matrix for the loadings (51).

Comment

If the data are complete or if the number of missing observations is the same for every test, $f_j^\star = 0$ for every j and the variance of $\sum_j^p f_j g_j^2$ is zero.

The variance formula (103c) can be considerably simplified when the data are complete. With a nonstochastic factor we obtain:

$$(105a) \quad D^2(u) = \frac{2}{k} (2p\alpha + 1)$$

$$\begin{aligned} (105b) \quad D^2(\hat{\alpha}_0) &= \frac{1}{(p-1)^3} \left[(p-1) D^2(u) + \frac{2(p\alpha + 1)^2}{k} \right] \\ &= \frac{2}{k(p-1)^3} \left[(p-1)(2p\alpha + 1) + (p\alpha + 1)^2 \right] \end{aligned}$$

and finally

$$\begin{aligned} (105c) \quad D^2(\hat{\alpha}) &= \frac{f^2}{p^2 k^2} D^2(\hat{\alpha}_0) \\ &= \frac{2}{k \cdot p^2 (p-1)} \left[(p-1)(2p\alpha + 1) + (p\alpha + 1)^2 \right] \end{aligned}$$

For a stochastic factor, we have:

$$(106a) \quad D^2(u) = \frac{2}{k} (p\alpha + 1)^2$$

$$(106b) \quad D^2(\hat{\alpha}_0) = \frac{2p}{k(p-1)^3} (p\alpha + 1)^2$$

and finally

$$(106c) \quad D^2(\hat{\alpha}) = \frac{2(p\alpha + 1)^2}{kp(p-1)}$$

The variance for $\hat{\alpha}$ can be estimated from the sample. When the data are complete we can use $\hat{\alpha}$ instead of α in formulae (105c) and (106c).

For incomplete data, the estimation of the variance of $\hat{\alpha}$ is somewhat more complicated. If the factor is nonstochastic, we can estimate the variance of u with

$$(107a) \quad \hat{D}^2(u) = \frac{2}{k} \left[2p\hat{\alpha} \sum_j^p f_j g_j^2 + 1 \right]$$

and if the factor is stochastic, with:

$$(107b) \quad \hat{D}^2(u) = \frac{2}{k} \left[2p\hat{\alpha} \sum_j^p f_j g_j^2 + 1 + (p\hat{\alpha})^2 \frac{1}{k} \sum_i^k \frac{1}{K_i^2} \right]$$

where $\hat{\alpha}$ in (107a,b) is given in (103).

The variance of $\hat{\alpha}_0$ is then estimated by

$$(108a) \quad \hat{D}^2(\hat{\alpha}_0) = \frac{k^2}{f^2} \left[\hat{D}^2(u) + \frac{2}{f} (p\hat{\alpha} \sum_j^p f_j g_j^2 + 1)^2 \right]$$

An estimate of $D^2(\sum_j^p f_j g_j^2)$ is

$$(108b) \quad \hat{D}^2(\sum_j^p f_j g_j^2) = 4 F^{\star'} \hat{\psi} F^{\star}$$

where $\hat{\psi}$ is the estimated covariance matrix for the loadings (84).

Then

$$(109) \quad \hat{D}^2(\hat{\alpha}) = \frac{1}{(\sum_j^p f_j g_j^2 \cdot pk)^2} \left[f \hat{D}(\hat{\alpha}_0) + pk\hat{\alpha} \hat{D}(\sum_j^p f_j g_j^2) \right]^2$$

is an estimate of the upper limit for the variance of $\hat{\alpha}$.

These results can be used to set up confidence intervals for the ratio α . Since the asymptotic distribution of $\hat{\alpha}_0$ is normal, it follows immediately that that of $\hat{\alpha}$ is also normal, and hence that an approximately 95% confidence interval for α is

$$(110) \quad \hat{\alpha} \pm 1.96 \hat{D}(\hat{\alpha})$$

Comments

Confidence intervals based on (110) will be conservative in two aspects:

i) When the proportions of missing observations are different for different tests, we expect the intervals to be too long because (109) and (103c) are estimates of the *upper* limits for the variance of $\hat{\alpha}$.

ii) Intervals corresponding to an $\hat{\alpha}$ larger than α will tend to be longer than those corresponding to an $\hat{\alpha}$ smaller than α . This follows directly from (107–109).

i) and ii) are clearly illustrated by a Monte Carlo experiment which is reported in section seven.

Using the above we obtain the following estimate of α_j ,

$$(111) \quad \hat{\alpha}_j = g_j^2 \cdot p \hat{\alpha}$$

which has asymptotic mean

$$(111a) \quad E(\hat{\alpha}_j) = p \gamma_j^2 \alpha = \alpha_j$$

For the asymptotic variance of $\hat{\alpha}_j$ we find

$$(111b) \quad D^2(\hat{\alpha}_j) \leq p^2 [\gamma_j^2 D(\hat{\alpha}) + 2\gamma_j \alpha D(g_j)]^2$$

This could be used to set up confidence intervals for α_j . To avoid the difficulties associated with the simultaneous estimation of a large number of confidence intervals (the estimated loadings are mutually correlated), it is better to use the limits for α and γ_j . For convenience, let the subscript L denote the lower limit of the interval. Then,

$$(112) \quad \alpha_{jL} = \gamma_{jL}^2 \cdot p \cdot \alpha_L$$

The upper limit of the confidence interval is obtained in the corresponding way.¹ These intervals are rather conservative as the correlation between g_j^2 and $\hat{\alpha}$ is probably small.

¹ Note that $0 \leq \gamma_j^2 \leq 1$. Thus if the confidence interval for γ_j contains the zero point, $\gamma_{jL}^2 = 0$.

Because of the one to one correspondence between α and $\bar{R}^2$ (and α_j and $\bar{R}_j^2$), we obtain the corresponding confidence intervals by substituting the appropriate confidence limits of α or α_j in (92a) and (92b).

6. THE ONE COMPONENT MODEL WITH UNKNOWN LOCATION PARAMETERS

The model (1) is sometimes written

$$(113) \quad \eta_1 = \nu_1 + \gamma_1 \xi + \epsilon_1$$

$$\eta_2 = \nu_2 + \gamma_2 \xi + \epsilon_2$$

.

.

$$\eta_p = \nu_p + \gamma_p \xi + \epsilon_p$$

where ν_j refers to unknown location parameters for the units of measurements used for η_j . These location parameters are often referred to as the mean values for the observed variables. For complete data, the location parameters are estimated by the mean values in the sample (Whittle, 1952), i.e. ν_j is estimated by

$$(114) \quad \bar{y}_j = \frac{1}{k} \sum_i^k y_{ij}$$

These mean values are subtracted from the observations, and the loadings and the component are estimated in the usual way.

6.1 Estimation when the data is incomplete

i) Approach I

We assume here that the missing observations are independent of the factor or component scores in the sense that the mean values of the factor scores for the missing observations are equal in every test. That is, we do not allow the missing observations for one test to correspond mainly to large (or to small) factor or component scores. If the above assumption is fulfilled, we can estimate the unknown parameters of the model in a way analogous to that for complete data, i.e. the location parameters are estimated by

$$(115) \quad \bar{y}_j = \frac{1}{N_j} \sum_i^k c_{ij} y_{ij}$$

where

$$(115a) \quad N_j = \sum_i^k c_{ij}$$

We set

$$(116) \quad y_{ij}^{\star} = y_{ij} - \bar{y}_j$$

and solve (15a-c) for the loadings and the component with $y_{ij}^{\star}$ in the iterative procedure (17a-c).

The convergence of the iterative procedure and the large sample properties of the MO method are still valid for the model written in (113), because the correction for the mean affects the consistency by an amount of order $\frac{1}{K}$ and the variance by terms of smaller order.

ii) Approach II

It is possible to obtain least square estimates of the unknown parameters in the model (113) without assuming that the missing observations are independent of the factor scores. We proceed as follows:

Writing the estimated model as

$$(117) \quad y_{ij} = v_j + g_j x_i + e_{ij}$$

with v_j as an estimate of the location parameter, we get the sum of squared residuals

$$(118) \quad q = \sum_i^k \sum_j^p c_{ij} (y_{ij} - v_j - g_j x_i)^2$$

Minimizing q subject to the condition $\sum_j^p g_j^2 = 1$ leads to

$$(119) \quad F = \sum_i^k \sum_j^p c_{ij} (y_{ij} - v_j - g_j x_i)^2 + \lambda (\sum_j^p g_j^2 - 1)$$

As before, we form the partial derivatives of F:

$$(119a) \quad \frac{\delta F}{\delta v_j} = -2 \sum_i^k c_{ij} (y_{ij} - v_j - g_j x_i)$$

$$(119b) \quad \frac{\delta F}{\delta g_j} = 2 \sum_i^k c_{ij} (y_{ij} - v_j - g_j x_i) (-x_i) + 2\lambda g_j$$

$$(119c) \quad \frac{\delta F}{\delta x_i} = 2 \sum_j^p c_{ij} (y_{ij} - v_j - g_j x_i) (-g_j)$$

Setting the partial derivatives equal to zero, we obtain

$$(120a) \quad \sum_i^k c_{ij} v_j = \sum_i^k c_{ij} (y_{ij} - g_j x_i)$$

$$(120b) \quad g_j (\sum_i^k c_{ij} x_i^2 - 2\lambda) = \sum_i^k c_{ij} (y_{ij} - v_j) x_i$$

$$(120c) \quad x_i \sum_j^p c_{ij} g_j^2 = \sum_j^p c_{ij} (y_{ij} - v_j) g_j$$

To show that the Lagrange multiplier is zero, we first define the variable y_{ij}^{**} by

$$(121) \quad y_{ij}^{**} = \begin{cases} y_{ij} & \text{if } c_{ij} = 1 \\ v_j + g_j x_i & \text{if } c_{ij} = 0 \end{cases}$$

Then, (120a–c) can be written

$$(122a) \quad \sum_j^p v_j = \sum_i^k (y_{ij}^{**} - g_j x_i)$$

$$(122b) \quad g_j \left(\sum_i^k x_i^2 - 2\lambda \right) = \sum_i^k (y_{ij}^{\star\star} - v_j) x_i$$

$$(122c) \quad x_i \sum_j^p g_j^2 = \sum_j^p (y_{ij}^{\star\star} - v_j) g_j$$

Multiplying (122b) by g_j and summing over j , we obtain

$$\begin{aligned} (123) \quad \sum_j^p g_j^2 \left(\sum_i^k x_i^2 + 2\lambda \right) &= \sum_j^p \sum_i^k (y_{ij}^{\star\star} - v_j) x_i g_j \\ &= \sum_i^k x_i \sum_j^p (y_{ij}^{\star\star} - v_j) g_j = \sum_i^k x_i^2 \end{aligned}$$

Thus, $\lambda = 0$.

The equations that define the estimates can now be rewritten as

$$(124a) \quad g_j = \frac{\sum_i^k c_{ij} (y_{ij} - \bar{v}_j) x_i}{\sum_i^k c_{ij} (x_i - \bar{x}_j)^2}$$

$$(124b) \quad \sum_j^p g_j^2 = 1$$

$$(124c) \quad v_j = \bar{y}_j - g_j \bar{x}_j$$

$$(124d) \quad x_i = \frac{\sum_j^p c_{ij} (y_{ij} - v_j) g_j}{\sum_j^p c_{ij} g_j^2}$$

where $\bar{y}_j$ is defined by (115), N_j by (115a) and $\bar{x}_j$ by

$$(124e) \quad \bar{x}_j = \frac{1}{N_j} \sum_i^k c_{ij} x_i$$

(It is assumed that $\sum_i^k c_{ij} (x_i - \bar{x}_j)^2$ and $\sum_j^p c_{ij} g_j^2$ are not zero.)

To obtain a unique solution of the system we must add another constraint such as requiring the arithmetic mean of the estimated component to be zero, i.e.

$$(124f) \quad \sum_i^k x_i = 0$$

To solve the system (124a–f) we can apply an iterative technique similar to (17a–c).

However, if there is no tendency for the missing observations to correspond mainly to large (or small) factor scores for different tests, $\bar{x}_j$ in (124e) will be very close to zero, and thus, $v_j \approx \bar{y}_j$. Then for all practical purposes it is of no consequence in large samples whether we simply estimate v_j by $\bar{y}_j$ or use the above least squares procedure.

On the other hand, if the actual observations correspond mainly to large (or small) factor scores, the denominator in (124a) will be small and the variance of the loadings will be large. An iterative procedure based on (124a–f) has been tried on a number of examples with negative results. If the missing observations were independent of the factor scores, approaches I and II gave practically identical results. On the other hand, if this was not the case, the procedure based on (124a–f) either failed to converge or converged to the wrong solution.

Because of this, the second approach has not been investigated further. The properties of the first approach are illustrated by a small Monte Carlo study in section seven.

7. MONTE CARLO EXPERIMENTS

To illustrate the large sample properties of the MO method, five small Monte Carlo studies have been carried out, using the CDC 3600 computer at UDAC. (Uppsala Data Central.)

Study 1

The results obtained in section four are illustrated with data generated according to the one factor model,

$$(125) \quad y_{ij} = \gamma_j \mu_i + \epsilon_{ij}$$

with a nonstochastic factor and normally distributed residuals. (The residuals were obtained by forming the sum of twelve independent rectangularly distributed variables.) The number of observations, k , was 100 and the number of variables, p , was 4.

Four models were tested:

- i) Model A: $\gamma_1 = .500$; $\gamma_2 = .500$; $\gamma_3 = .500$; $\gamma_4 = .500$; $\sigma_\mu^2 = 9$
and $\sigma^2 = 4$
- ii) Model B: $\gamma_1 = .500$; $\gamma_2 = .500$; $\gamma_3 = .500$; $\gamma_4 = .500$; $\sigma_\mu^2 = 16$
and $\sigma^2 = 4$
- iii) Model C: $\gamma_1 = \frac{1}{\sqrt{10}}$; $\gamma_2 = \frac{2}{\sqrt{10}}$; $\gamma_3 = \frac{-1}{\sqrt{10}}$; $\gamma_4 = \frac{-2}{\sqrt{10}}$;
 $\sigma_\mu^2 = 9$ and $\sigma^2 = 4$
- iv) Model D: $\gamma_1 = \frac{1}{\sqrt{10}}$; $\gamma_2 = \frac{2}{\sqrt{10}}$; $\gamma_3 = \frac{-1}{\sqrt{10}}$; $\gamma_4 = \frac{-2}{\sqrt{10}}$;
 $\sigma_\mu^2 = 16$ and $\sigma^2 = 4$

These models may appear unrealistic in the sense that the variance of the factor is large compared to the residual variance, but, in reality, the opposite holds. The factor accounts for 50% of the variability in models B and D,

¹ The loadings of this model are the same as those used by Wold (1966b).

while for models A and C, only 36%. In practice, the factor in the one factor model will account for about 50% or more.

Ten different proportions of missing observations were generated randomly according to:

TABLE 1. *Types of Incompleteness*

% missing in variable	T Y P E									
	I	II	III	IV	V	VI	VII	VIII	IX	X
1	0	10	20	30	40	50	0	10	20	30
2	0	10	20	30	40	50	10	20	30	40
3	0	10	20	30	40	50	20	30	40	50
4	0	10	20	30	40	50	30	40	50	60

For each model, the experiment was performed so that

- a) The factor was the same for all types of incompleteness.
- b) The positions of the missing observations were constant for every type of incompleteness.

For each type of incompleteness 100 samples were generated and the following calculated:

- 1) The arithmetic means for the loadings.
- 2) The sample covariance matrix for the loadings.
- 3) The covariance matrix for the loadings according to (51).
- 4) The arithmetic mean and variance for the estimated residual variance; these values can be compared with the corresponding asymptotic quantities:

$$E(\hat{\sigma}^2) = \sigma^2 \quad \text{and} \quad D^2(\hat{\sigma}^2) = \frac{2 \sigma^4}{f^{\star}}$$

- 5) The number of times that the confidence intervals based on the covariance matrices (51) and (84)¹ failed to cover the true loadings.
- 6) A chi-square test (with 6 degrees of freedom) based on the likelihood

¹ The estimated covariance matrix is based on those observations ("i") for which the denominator in (81a) is greater than or equal to 0.0001.

ratio criteria for testing the hypothesis that the covariance matrix for the loadings is that given by (51). (See for example Anderson, 1958.)

The results from this Monte Carlo study are given in the appendix, Tables A – D.

Comments

As the theoretical covariance matrix is singular, the last row and the last column are eliminated when the chi-square test is performed. For samples from a normal distribution the test variable is asymptotically chi-square distributed. As the loadings are asymptotically normally distributed with covariance matrix equal to ψ (51) it follows that we can regard this test only as an approximate measure of the deviations between the theoretical and the observed covariance matrices.

From the tables we see that the estimated and theoretical covariance matrices for the loadings agree rather well, even if the proportion of missing observations is large. The same is true for the arithmetic means of the loadings and the estimated residual variance. This evidence suggests that even when there is a large proportion of missing observations, the bias for the MO method is small.

We also observe that the confidence intervals based on the theoretical and on the estimated covariance matrices agree reasonably well. Thus, the large sample properties of the MO method derived in section four seem to be valid even for samples of moderate size.

Study 2

In this study we shall illustrate the properties of the MO method for the one factor model with unknown location parameters by using model D in Study 1. With the location parameters set equal to 1, the model becomes

$$(126) \quad y_{ij} = 1 + \gamma_j \mu_i + \epsilon_{ij}$$

The random numbers used to generate the data for this study were identical with those used for model D in the first study. The results from the two studies are therefore directly comparable. To estimate the parameters of the model we have used the first approach of section six. The mean value of the

corresponding variables are subtracted from each observation and the parameters are estimated with the MO method. The results of this study are shown in Table E of the appendix.

Comments

It should be noted that, because we subtract the arithmetic means from the observations, the actual values of the location parameters are irrelevant to the outcome of the experiment.

If the missing observations do not correspond to mainly large or small factor scores, the properties of the MO method remain the same whether or not we know the location parameters. This can be seen in the similarities between Tables E and D as was expected from the effects caused by the elimination of the mean values (section 6.1). The only notable differences are observed for incompleteness of types VI and X, i.e. for cases where the proportion of missing observations are large. It should also be noted that the arithmetic mean of the estimated residual variance is always smaller in Table E than in Table D. This arises because in the second study we have again used $f^{\star}$ for the degrees of freedom of the estimated residual variance. But since we have, in fact, estimated another p quantities (the location parameters), we should use $f^{\star} - p$ for the degrees of freedom. (As $f^{\star}$ is usually much larger than p , the effect of using $f^{\star}$ instead of $f^{\star} - p$ will be small; this is shown in the tables.)

Study 3

To illustrate the bias in the estimates of the loadings when the missing observations are supplemented by the mean values of the corresponding tests, we have chosen the incompleteness of type VII in model B of Study 1.

100 samples were generated and the results are shown in Table 2.

From the table we see that the bias for this method is considerably larger than for the MO method. The relative bias is between 2 and 24%, while for the MO method it was less than 2%. (Note that the standard deviations for the mean values are about 0.005 – 0.006.)

TABLE 2. *Loadings, means and variances. 100 samples.*

Variable	% missing	Mean value	Variances
1	0	.553	2.92 10^{-3}
2	10	.547	2.82 10^{-3}
3	20	.491	2.41 10^{-3}
4	30	.379	2.17 10^{-3}

Study 4

The results of section four, are valid for both stochastic and nonstochastic factor. We have used model D of the first study to demonstrate this fact. In the model, the previous non stochastic factor has been replaced by a stochastic factor which is normally distributed with a mean of zero and variance equal to 16. The results of this study are recorded in Table F of the appendix where we see that they are similar to those of the corresponding model in Study 1.

Study 5

In this study we consider the behaviour of $\hat{\alpha}$, i.e., the estimate of the ratio between explained and unexplained variation. As in Study 4, we have chosen model D of Study 1, in which the factor accounts for 50% of the variation. Thus α is equal to 1.

For each type of incompleteness and for both types of factor (stochastic and non-stochastic), we generated 100 samples. In the nonstochastic case we have set all factor scores equal to 4 in order that assumption (96) should be fulfilled.

The results are shown in the following tables:

TABLE 3 a. Stochastic factor. $\alpha = 1$. $k = 100$.

Charac- teristic	TYPE OF INCOMPLETENESS									
	I	II	III	IV	V	VI	VII	VIII	IX	X
(1)	1.01	1.06	1.07	1.07	1.08	1.09	1.08	1.06	1.06	1.10
(2)	.204	.222	.245	.275	.316	.377	.290	.331	.388	.475
(3)	.214	.229	.242	.321	.263	.335	.221	.285	.273	.329
(4)	5	9	8	8	3	2	3	3	3	3
(5)	4	9	8	8	3	2	3	3	3	0
(6)	4	2	1	6	0	3	2	1	0	0
(7)	2	0	0	2	0	0	0	0	0	0

where the contents of the rows are as follows:

- (1) Arithmetic mean of $\hat{\alpha}$ in the 100 samples.
- (2) Theoretical standard deviation of $\hat{\alpha}$ (see 103c).
- (3) Estimated standard deviation of $\hat{\alpha}$ in the 100 samples.
- (4) Number of 95% confidence intervals based on (103c) which do not cover α .
- (5) Number of 95% confidence intervals based on (103c) with lower limit larger than α .
- (6) Number of 95% confidence intervals based on (109) not covering α .
- (7) Number of 95% confidence intervals based on (109) with lower limit larger than α .

TABLE 3 b. Nonstochastic factor. $\alpha = 1$. $k = 100$.

Charac- teristic	TYPE OF INCOMPLETENESS									
	I	II	III	IV	V	VI	VII	VIII	IX	X
(1)	1.05	1.05	1.06	1.07	1.10	1.13	1.03	1.05	1.07	1.10
(2)	.147	.159	.174	.193	.221	.265	.217	.248	.292	.357
(3)	.149	.159	.168	.167	.221	.300	.167	.169	.213	.291
(4)	5	7	6	5	9	11	2	2	1	6
(5)	5	7	6	5	9	11	2	2	1	6
(6)	3	2	3	3	2	4	0	1	1	0
(7)	3	1	2	1	0	3	0	1	0	0

where the characteristics are the same as in Table 3a.

Comments

Incompleteness of types VII–X correspond to cases where the proportions of missing observations are different for different tests. According to section five we then expect the estimated standard deviations to be smaller than the theoretical. Further we expect rather few "significant" confidence intervals. Tables 3a–b confirm these expectations.

We see also that the observed distributions of $\hat{\alpha}$ are not symmetric because most of the "significant" intervals based on the theoretical standard deviations have midpoint larger than α .

We would also expect that there would not be many "significant" confidence intervals with a midpoint greater than α which are based on the estimated parameters. This is borne out by Tables 3a–b in which 96.2% of the "significant" intervals based on the true parameters, have midpoint greater than α whereas for intervals of this type based on the estimated parameters, the percentage is only 39.5.

8. APPLICATIONS

As stated by Wold (1966a) there are many situations in applied research where component analysis has to cope with incomplete data. In this section we shall give two applications, both from the field of education.

Application 1. Knowledge of geography at elementary school level.

The first application is given mainly for illustrative purposes. It is based on a rather small sample, and therefore, it cannot be expected that the large sample properties of the MO method will hold. However, in the present context, its small size is advantageous, since it enables us to display the fundamentals of the method without having to present a large quantity of results.

The example originates from a survey performed by Koort (Koort and Barrefors (1966), Koort (1969)). In his study, 16 geographical tests¹ were given to 45 elementary school classes. The data were highly incomplete; of the 720 possible observations there were only 283. This incompleteness is not unreasonable; it is obvious that it would be meaningless to give all tests to every class, irrespective of what the teacher has taught.

In order to obtain correspondence between courses and tests, each teacher chose the appropriate fraction of the 16 tests. (See Koort *et al.*, (1961), Koort (1969).)

Our model is the one factor model with all residual variances equal:

$$(127) \quad y_{ij} = \nu_j + \gamma_j \mu_i + \epsilon_{ij}$$

In this case the location parameters, ν_j , are known, and, in fact are equal to 3 for all j . Thus the estimated model is

$$(128) \quad y_{ij} - 3 = g_j x_i + e_{ij}$$

The data are shown in Table 4.

¹ The tests are PP2, PP4-PP18 in Koort and Barrefors (1966).

We shall use the MO method for exploratory purposes and try to find out if the factor accounts for 50% or more of the variability for the different tests. Before proceeding, we first note that the device of excluding all individuals (in this case, classes) who have not participated in every test results in no data at all left for analysis.

When the MO method was applied to the original set of data, test 8 showed an extremely high loading, about three times larger than those for the other tests. As it was given to only three classes, it was excluded and the data reanalysed. The loadings, the standard deviations for the loadings and the component shown in Table 4 refer to the reduced set of data.

We can now easily compute the confidence intervals for the loadings. From (87b) and (84) we find the interval

$$(129) \quad g_j \pm \sqrt{\chi^2_{1-\alpha} (p-1) \hat{\phi}_{jj}}$$

For γ_1 , we find the following interval (95% level):

$$(130a) \quad 0.17 \pm \sqrt{23.7} \cdot 0.043$$

or

$$(130b) \quad 0.17 \pm 0.209$$

This interval covers the zero point and implies that the loading for the first test does not differ significantly from zero at the 5% level. The same is true for tests 6, 13, 15 and 16. The loadings that differ significantly from zero at the 5% level are marked with an ★ in Table 4.

For this data set, we obtain the following estimate of α ,

$$(131a) \quad \hat{\alpha} = 3.07$$

The estimated standard deviation of $\hat{\alpha}$ is 2.28 for a stochastic factor and 1.87 for a nonstochastic factor, both of which lead to confidence intervals (level 95%) containing the zero point. These intervals are probably too long as the estimated standard deviations are estimates of upper limits for the true standard deviations.

TABLE 4. Class means and results of the component analysis.

[illegible]

Using the above, we obtain the estimate of $\bar{R}^2$.

$$(131b) \quad R^2 = \frac{\hat{\alpha}}{1 - \hat{\alpha}} = .75$$

However, in this example we are mainly interested in the amount of variability in each test accounted for by the factor. For each test we have the following estimate of the ratio between explained and unexplained variation,

$$(132a) \quad \hat{\alpha}_j = g_j^2 \hat{\alpha} p$$

and the estimate of the ratio between explained and total variability,

$$(132b) \quad R_j^2 = \frac{\hat{\alpha}_j}{1 + \hat{\alpha}_j}$$

For example, in the first test we have

$$(133a) \quad \hat{\alpha}_1 = 0.17^2 \cdot 3.07 \cdot 15 = 1.33$$

and

$$(133b) \quad R_1^2 = \frac{1.33}{2.33} = 0.57$$

This gives a correlation between the first test and the factor equal to 0.75.

Because the confidence interval for α contains the zero point, the intervals for α_j and $\bar{R}^2$ will also contain the zero point.

If we examine the results more closely, we find that test 16 shows some peculiarities. First, it has a very high standard deviation for the loading. Figure 1 shows its observations plotted against the component and its regression on the component. We see that the variability in the component for test 16 is quite small. To a large extent, this explains the size of the standard deviation for test 16. Secondly, the residuals are all positive. However, as there are only five observations, it is not too unlikely that this could occur by chance. Another possibility is that the location parameter for test 16 is not equal to 3. For purposes of comparison, we have plotted the same quantities for test

6, and in this case we see that the variability of the component is reasonably large and that the regression line passes through the observations instead of lying below them.

In conclusion, it seems likely from this exploratory study that the tests, excluding test 8 (and perhaps also test 16), measure the same latent variable, which could be called "general knowledge of geography". The estimate of the correlation between factor and test is rather high (greater than 0.7 for each test). However, to be able to make a more precise statement we would need a larger sample.

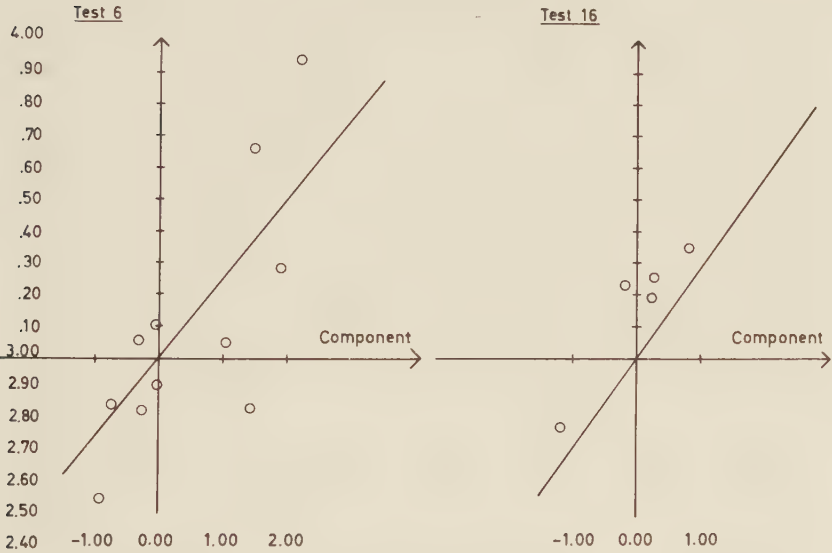


Fig. 1. Test 6 and test 16 plotted against the estimated component.

Application 2. Duration of studies.

This application is taken from a study about the length of studies for one and two points (betyg) in different subjects at the faculties of social sciences in Sweden for the academic year 1964/65 and 1965/66 (Christoffersson, 1970a).

Since the academic year 1963/64, the Swedish National Central Bureau of Statistics (SCB) has been collecting data about the length of studies in all subjects in the faculties of arts and sciences. These data are based on a form (Tentamensdeklaration), which each student has to complete in connection with the examination. One of the purposes of these surveys is to "... enable checks on the extent in which the students of different subjects pass their examination in normal time. . . (SCB, 1967a)". "Normal time" means a net length of studies of five months for one point and ten months for two points. The net length is defined as the time between the beginning of the studies in the subject and the passing of the examination with a "deduction for holidays, full- or part-time economical activities, illness and other things that may have a negative effect on the possibility to study. . . (SCB, 1967a)". With this definition, the length of studies for two points will include the time for one point.

In the surveys published by SCB on this subject, no causal element has been introduced. This implies that the main use to which the results of these surveys can be put is to find out whether the net length of studies is too long or not. As shown by Christoffersson (1970a), the one factor model makes it possible to introduce more differentiation into the picture of the causal complex that affects the length of studies. This in turn, suggests where resources should be allocated in order to diminish the problem of long net lengths of studies.

We have the following model

$$(134) \quad NT_j - y_{ij} = \gamma_j \mu_i + \epsilon_{ij} \quad j = 1, 2 \dots p; i = 1, 2 \dots k$$

where NT_j is the normal time for the j :th subject, y_{ij} denotes the net length of studies for the i :th student in the j :th subject, γ_j refers to the degree of difficulty for the j :th subject, μ_i to the capacity for the i :th student and ϵ_{ij} is the residual. In (134) one and two points in the same subject are treated as two different "subjects". It is also assumed that the residual variance for two points is twice the residual variance for one point.¹

In this way we obtain a decomposition of the net length of studies into two causal "factors", one referring to subject and the other to the individual student. High capacity for a student is manifested by short net length of studies and corresponds to high positive values of the capacity factor μ_i .

¹ For a more detailed discussion of the applicability of the model, see Christoffersson (1970a).

Instead of using γ_j as a measure of the difficulty of a subject, we shall use a quantity D_j for which the interpretation is rather easier. D_j is defined to be the expected net length of studies in subject j for a student of average capacity, i.e.

$$(135) \quad D_j = NT_j - \gamma_j \bar{\mu}$$

where $\bar{\mu} = \frac{1}{k} \sum_{i=1}^k \mu_i$ is the average capacity.

For the sake of brevity, we shall consider only the faculty of social sciences at the University of Uppsala. We shall divide the analysis into three steps. The analysis is based on SCB's primary material of net length of studies for one and two points during the academic years 1964/65 and 1965/66. In the first step of the analysis, those students who have passed only one examination during the two years were excluded because they do not affect the estimates of the γ_j 's. The subject human geography has also been excluded because in SCB's material, the two subjects human geography and geography have been recorded in the same way.

As no empirical evidence about the quality of the data has been given by SCB, the second step of the analysis was a simple check of the reasonableness of the data. The net length of studies for two points consists of net length of studies for one point plus net length of studies for the two point course. For one and the same student, the net length of studies for two points in a subject should be longer than the corresponding time for one point in the same subject. This is not always the case in SCB's material as can be seen from Table 5.

As can be seen from the table, SCB's primary material has low quality. The question might well be asked whether it is possible to make a meaningful analysis using such material. However, after having removed "absurd" data according to the following rules, it should be possible to extract useful information from the remaining material.

- a) Data for which the net length of studies for one point is less than two months or greater than 15 months are excluded.
- b) Data for which the net length of studies for two points is less than four months or greater than 30 months are excluded.

TABLE 5. *Percentage of students at the faculty of social science at the University of Uppsala who, according to SCB's material for length of studies, passed the examination for two points in a shorter time than they passed the examination for one point in one and the same subject. (Refers to those who during the period passed the examinations for one and two points in at least one subject.)*

Subject	Number of students who passed the examinations for one and two points	Percentage that passed the examination for two points in less time than one point
Economic History	3	0.0
Economy of the firm	63	12.7
Economics	59	15.3
Pedagogics	29	17.2
Psychology	70	28.6
Sociology	106	30.2
Statistics	121	16.5
Political Science	102	11.8
Total	553	19.2

c) If the net lengths of studies for two points is less than the time for one point plus two months, the observations for one and two points are excluded.

This process will reduce the material by about 20%.

In the data there is a large proportion of missing observations (85.4%). These missing observations can be divided in two categories. The first category are missing observations due to the process of excluding "absurd" data or because the student did not fill in the form (Tentamensdeklaration) correctly (see SCB, 1967a). The second category are observations that are missing because a student does not study every subject at the faculty. Naturally, any results obtained from these data must be interpreted with great care.

The third step of the analysis is the estimation of the unknowns in the model (134). The estimated model is

$$(136) \quad NT_j - y_{ij} = g_j x_i + e_{ij}$$

Because the residual variance for two points is assumed to be twice the variance for one point we introduce

$$(137a) \quad y_{ij}^{\star} = NT_j - y_{ij} \quad \text{if } j \text{ corresponds to a one point subject}$$

$$(137b) \quad y_{ij}^{\star} = \frac{NT_j - y_{ij}}{\sqrt{2}} \quad \text{if } j \text{ corresponds to a two points subject}$$

and rewrite the estimated model

$$(138) \quad y_{ij}^{\star} = g_j^{\star} x_i + e_{ij}^{\star}$$

We now use the MO method to estimate $g_j^{\star}$ and x_i , using $y_{ij}^{\star}$ and $g_j^{\star}$ in (17a–c) instead of y_{ij} and g_j . The final estimates of γ_j are given by

$$(139a) \quad g_j = g_j^{\star} \quad \text{if } j \text{ corresponds to one point}$$

$$(139b) \quad g_j = \sqrt{2} g_j^{\star} \quad \text{if } j \text{ corresponds to two points}$$

According to (15b) we obtain the estimated capacity as

$$(139c) \quad x_i = \frac{\sum_j^p c_{ij} y_{ij}^{\star} g_j^{\star}}{\sum_j^p c_{ij} (g_j^{\star})^2}$$

From (139c) we see that the numerical value of x_i can be very large if the denominator of (139c) is small which implies that the variance of x_i is large if the denominator is small. Therefore, the average capacity $\bar{\mu}$ is estimated with the arithmetic average of the estimated capacity over those i s for

which the denominator in (139c) is larger than or equal to 0.01. Denoting this average with $\bar{x}$, we estimate the expected net length of studies in the j :th subject for a student with average capacity by

$$(140) \quad d_j = NT_j - g_j \bar{x}$$

The results of the analysis are shown in Table 6, together with the net length of studies according to SCB. The estimated capacity is shown in Figure 2.

The estimate of the ratio of explained to unexplained variation is

$$(141a) \quad \hat{\alpha} = 0.819$$

and its estimated standard deviation is (for nonstochastic factor)

$$(141b) \quad \hat{D}(\hat{\alpha}) = 0.283$$

As mentioned in section five, $\hat{D}(\hat{\alpha})$ is an estimate of an upper limit for the standard deviation of $\hat{\alpha}$. Thus, the standard deviation of $\hat{\alpha}$ is probably less than 0.283. Using the above we find a 95% confidence interval for α to be

$$(141c) \quad 0.819 \pm 0.555$$

An estimate of the ratio of explained to total variability is

$$(142a) \quad R^2 = \frac{\hat{\alpha}}{1 + \hat{\alpha}} = 0.450$$

i.e. the factor accounts for about 45% of the total variability in the material. A 95% confidence interval for this ratio is

$$(142b) \quad 0.209 \text{ ————— } 0.579$$

Due to the quality of the primary material, the results must be interpreted with great care. However, the results do give indications about where to look

TABLE 6. *Net length of studies in the social science faculty at the University of Uppsala. Academic year 1964/65 and 1965/66.*

Subject	Point (betyg)	SCB 1964/65 Median	SCB 1965/66 Median	Arith- metic mean 1964/65- 1965/66	d_j	g_j	$(\hat{\psi}_{jj})^{1/2}$	Number of students
Economic history	1	4	4	6	5	.00	.05	20
	2	10	11	11	11	.15	.08	46
Economy of the firm	1	5	5	6	6	.09	.03	171
	2	12	11	12	11	.24	.06	88
Economics	1	6	6	6	6	.22	.04	199
	2	16	14	16	14	.67	.07	121
Pedagogics	1	4	5	5	5	.03	.02	232
	2	6	5	10	11	.09	.08	52
Psychology	1	4	—	5	5	.02	.06	57
	2	9	10	11	10	.03	.03	72
Sociology	1	5	5	6	5	.06	.03	346
	2	8	11	12	13	.59	.10	120
Statistics	1	4	4	5	6	.13	.05	306
	2	8	9	10	12	.29	.07	139
Political science	1	6	7	7	6	.23	.03	330
	2	13	14	15	15	.86	.09	168

d_j = Expected net length of studies.

g_j = Degree of difficulty.

$(\hat{\psi}_{jj})^{1/2}$ = Standard deviation of g_j .

The arithmetic mean is based on the results after they have been screened as described in the second step of the analysis. The effect of this screening is most marked for Pedagogics (two points).

Average study capacity, $\bar{x} = -5.3$

Number of students = 1059

SCB supplied no estimate for Psychology, one point, in 1965/66 as the number of students was less than 10.

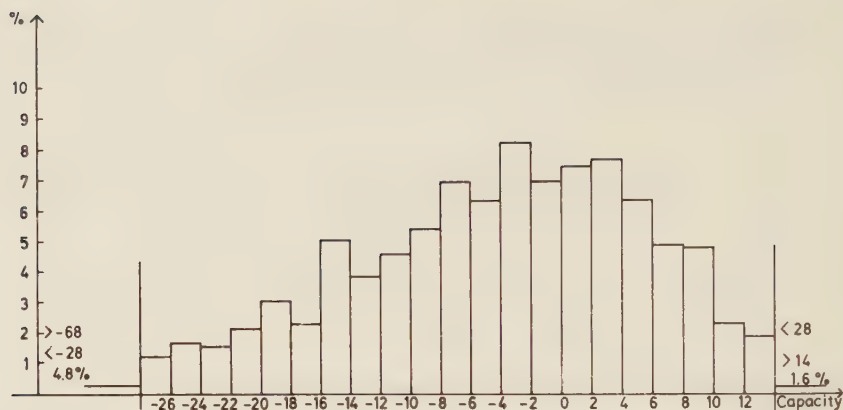


Fig. 2. Estimated capacity. (Based on results for the 868 students for whom the denominator in (139c) is greater than or equal to 0.01.)

for the causes of long net lengths of studies. In economics (two points) the average net length of studies is 16 months. As the expected net length of studies is 14 months, the main causes for the long net length of studies relate to the subject whereas the difference between the average and the expected net length of studies, 2 months, refers to the students. That is, the average capacity of students in economics seems to be lower than for the faculty as a whole. In statistics (two points), the effect of the difficulty of the subject seems to have been completely eliminated by the high capacity of the students who choose to take this subject. For political science we find that the average net length of studies and expected net length are equal, thus indicating that the causes of the long net length of studies relate to the subject itself and not to the students.

Comments

The computing time for the first application was 6 seconds and for the second 4 minutes and 52 seconds, using the CDC 3600 computer at UDAC. For both applications the first approximation for the loadings was $g_j = 1$. In the first application the number of iterations was 15, using $|g_j^{(s)} - g_j^{(s+1)}| < 10^{-6}$ as convergence criteria. In the second application the process converged after 137 iterations using the criteria $|g_j^{(s)} - g_j^{(s+1)}| < 10^{-4}$.

9. THE MULTIFACTOR (COMPONENT) MODEL

The main theme of this monograph is the problem of missing observations in a one factor or a one component model. In this section, we depart from this theme in order to present a brief discussion of the multifactor model. If there is more than one factor (component), the model (1) takes the following form

$$\begin{aligned}
 (143) \quad \eta_1 &= \gamma_{11} \xi_1 + \gamma_{12} \xi_2 + \dots + \gamma_{1n} \xi_n + \epsilon_1 \\
 &\cdot \\
 &\cdot \\
 \eta_p &= \gamma_{p1} \xi_1 + \gamma_{p2} \xi_2 + \dots + \gamma_{pn} \xi_n + \epsilon_p
 \end{aligned}$$

where the residuals $\epsilon_1, \epsilon_2, \dots, \epsilon_p$ are assumed to have zero expectation and no correlation with the latent variables $\xi_1, \dots, \xi_n$. Further, the loadings $\gamma_{11}, \dots, \gamma_{pn}$ are normalized so that

$$(143a) \quad \sum_j^p \gamma_{jr}^2 = 1 \quad \text{for } r = 1, 2, \dots, n$$

If we assume that

$$(144) \quad \sum_j^p \gamma_{jr} \epsilon_j = 0 \quad \text{for } r = 1, 2, \dots, n$$

we obtain the component representation of the model. From (144) it follows that the components are linear in the observed variables.

On the other hand, if we assume that the residuals are mutually uncorrelated, we get the factor representation of the model. These representations of the model are unique only up to a nonsingular transformation of the matrix formed by the loadings.

In the component representation it is common to assume that the components $\xi_1, \xi_2, \dots, \xi_n$ are mutually orthogonal and that ξ_1 is the linear combi-

nation of $\eta_1, \eta_2, \dots, \eta_p$ which has the largest possible variance, that ξ_2 is the linear combination which is uncorrelated with ξ_1 and has the largest possible variance, and so on. With these restrictions (which only determine a set of basis vectors in the space spanned by the components) the component representation will be unique. If we assume that the residuals in the factor representation of the model have equal variances, the loadings are the same in both representations (Lyttekens, 1966). Under the assumption of normally distributed residuals, the eigenvectors of the sample covariance matrix, corresponding to the n largest eigenvalues, will be maximum likelihood estimates of the loadings. The maximum likelihood estimates of the components, or the factors corresponding to the i th observation, are linear combinations of the observation $y_{i1}, y_{i2}, \dots, y_{ip}$ with the elements of the eigenvectors as weights. (It should be noted that the method of least squares would also give the same estimates.) The numerical values of these estimates can be obtained by Wold's iterative procedure (7a-c). Using this procedure we first get estimates of the first component (factor) and the corresponding loadings. If we now substitute $(y_{ij} - g_j x_i)$ for y_{ij} in (7a-c) and repeat the process we obtain the estimates of the second component (factor) and the corresponding loadings, and so on.

The above procedure is valid if the data are complete. If the data are incomplete, the situation is more complicated. To simplify the discussion, let us consider the two-factor (component) model. The estimated model is written

$$(145) \quad y_{ij} = g_j x_i + h_j z_i + e_{ij}$$

The unknowns in this model may be estimated by a two-step procedure. In the first step, g_j and x_i are estimated using the iterative procedure (17a-c). Substituting $(y_{ij} - g_j x_i)$ for y_{ij} , h_j for g_j and z_i for x_i and repeating the process we obtain estimates of the second factor (component) and the corresponding loadings.

The estimates obtained in this way will, in general, not be the least square estimates. The sum of squared residuals is

$$(146) \quad f = \sum_i^k \sum_j^p c_{ij} (y_{ij} - g_j x_i - h_j z_i)^2$$

If this sum is minimized, we obtain the following system of non-linear equations

$$(146a) \quad g_j = \frac{\sum_i^k c_{ij} y_{ij} x_i}{\sum_i^k c_{ij} x_i^2} - h_j \frac{\sum_i^k c_{ij} z_i x_i}{\sum_i^k c_{ij} x_i^2}$$

$$(146b) \quad h_j = \frac{\sum_i^k c_{ij} y_{ij} z_i}{\sum_i^k c_{ij} z_i^2} - g_j \frac{\sum_i^k c_{ij} x_i z_i}{\sum_i^k c_{ij} z_i^2}$$

$$(146c) \quad x_i = \frac{\sum_j^p c_{ij} y_{ij} g_j}{\sum_j^p c_{ij} g_j^2} - z_i \frac{\sum_j^p c_{ij} h_j g_j}{\sum_j^p c_{ij} g_j^2}$$

$$(146d) \quad z_i = \frac{\sum_j^p c_{ij} y_{ij} h_j}{\sum_j^p c_{ij} h_j^2} - x_i \frac{\sum_j^p c_{ij} g_j h_j}{\sum_j^p c_{ij} h_j^2}$$

This system might be solved by an iterative process similar to that used for the one factor (component) model, i.e. for given $x_i^{(s)}$ and $z_i^{(s)}$ we solve for $g_j^{(s+1)}$ and $h_j^{(s+1)}$ in (146a-b) and then for given $g_j^{(s+1)}$ and $h_j^{(s+1)}$ we obtain $x_i^{(s+1)}$ and $z_i^{(s+1)}$ from (146c-d) and so on. The convergence properties of such an iterative procedure and the stochastic properties of the estimates, have been treated by Christofferson (1970b), who used the same approach as in section four of this monograph.¹

¹ M. Hagnell (1970) has been working on this problem from the point of view of component analysis with reference variables (Wold, 1966a). Hagnell has derived the second order moments for the loadings utilizing the approach of section four in this monograph. His results are similar to those in Christofferson (1970b).

If there are no missing observations in some of the tests, we might, as mentioned in the introduction, use one of the methods given in Horst (1965). Before discussing these methods it will be convenient to introduce some new notation.

Let

$$(147a) \quad \eta = \begin{pmatrix} \eta_1 \\ . \\ . \\ \eta_p \end{pmatrix}$$

$$(147b) \quad \Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & . & . & \gamma_{1n} \\ . & & & & \\ . & & & & \\ \gamma_{p1} & . & . & . & \gamma_{pn} \end{pmatrix}$$

$$(147c) \quad \xi = \begin{pmatrix} \xi_1 \\ . \\ . \\ \xi_n \end{pmatrix}$$

$$(147d) \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ . \\ . \\ \epsilon_p \end{pmatrix}$$

Using the above, we can write the model (143) as

$$(148) \quad \eta = \Gamma \xi + \epsilon$$

Further, let

$$(149a) \quad E(\eta \eta') = \Sigma$$

$$(149b) \quad E(\xi \xi') = \Lambda$$

$$(149c) \quad E(\epsilon \epsilon') = \Omega$$

For the factor representation of the model, Ω will be diagonal and Σ can be expressed as follows

$$(150) \quad \Sigma = \Gamma \Lambda \Gamma' + \Omega$$

If Λ is diagonal, i.e. if the latent factors are orthogonal, this decomposition will be unique up to an orthogonal transformation. Further, if we assume that

$$(151) \quad \Omega = \theta I$$

where θ is a constant, Γ will be the same for both the factor and the component representation (Lytkens, 1966).

If we assume that the variables η are standardized, i.e. they have zero mean and unit variance, Σ will be the correlation matrix for η . The loadings of the model can then be estimated by the eigenvectors of the sample correlation matrix which correspond to the k largest eigenvalues. These estimates will be consistent (Anderson, 1963). As the size of the sample increases, the sample eigenvalues will tend to

$$(152) \quad \Lambda + \theta I$$

Thus, the eigenvalues in the sample will not be consistent estimates of the

factor variances. (In passing, we note that the "best" estimates of the parameters in the model, under the assumption of normal distribution, are given by the maximum likelihood method (Lawley, 1940) which does not assume that the residual variances are equal.)

Let us consider the case in which the variables $\eta_1, \eta_2, \dots, \eta_p$ can be divided into two sets, such that every factor is present in at least two variables in each set. Then we can rewrite the model (148) as

$$(153) \quad \begin{bmatrix} \eta_\alpha \\ \eta_\beta \end{bmatrix} = \begin{bmatrix} \Gamma_\alpha \\ \Gamma_\beta \end{bmatrix} \xi + \begin{bmatrix} \epsilon_\alpha \\ \epsilon_\beta \end{bmatrix}$$

where η_α denotes the variables in the α set, Γ_α are the loading and ϵ_α the residual for the α set etc.

Corresponding to 150 we obtain

$$(154) \quad \Sigma = \begin{bmatrix} \Sigma_{\alpha\alpha} & \Sigma_{\alpha\beta} \\ \Sigma_{\beta\alpha} & \Sigma_{\beta\beta} \end{bmatrix} = \begin{bmatrix} \Gamma_\alpha \Lambda \Gamma'_\alpha & \Gamma_\alpha \Lambda \Gamma'_\beta \\ \Gamma_\beta \Lambda \Gamma'_\alpha & \Gamma_\beta \Lambda \Gamma'_\beta \end{bmatrix} + \begin{bmatrix} \Omega_\alpha & 0 \\ 0 & \Omega_\beta \end{bmatrix}$$

For the sample correlation matrix R we get

$$(155) \quad R = \begin{bmatrix} R_{\alpha\alpha} & R_{\alpha\beta} \\ R_{\beta\alpha} & R_{\beta\beta} \end{bmatrix}$$

where $R_{\alpha\alpha}$ and $R_{\beta\beta}$ denote the estimated correlations within the α set and the β set respectively. $R_{\alpha\beta} = R'_{\beta\alpha}$ denotes the estimated intercorrelations between the two sets.

The methods described in Horst (1965) are based on this decomposition and are applicable to situations in which there are missing observations only for the β set. Suppose that we have observations on the variables in the two sets for a group of individuals and that there is a large proportion of missing

observations for the β set but no missing for the α set. Then we can obtain $R_{\alpha\alpha}$ in the usual way. An element in $R_{\alpha\beta}$ is the correlation between one variable in the α -set and one in the β -set. This correlation can be approximated by the correlation between the two variables for the individuals for which we have observed the variable in the β set. (We could estimate the elements of $R_{\beta\beta}$ in this way but some (or all) of the correlations might be based on too few observations.) We now wish to find a method for estimating the parameters of the model which does not use these elements of $R_{\beta\beta}$.

The first method in Horst (1965) is the inter-battery method of factor analysis (Tucker, 1958). It utilizes only $R_{\alpha\beta}$ and $R_{\beta\alpha}$. However, as shown by Gibson (1960, 1961), there is a fundamental indeterminacy in the method which "cast grave doubt upon the general applicability of the inter-battery method in its original form" (Gibson, 1960).¹ For this reason, we shall not discuss the method further but turn instead to Horst's second and third methods.

In the second method, $R_{\alpha\alpha}$ and $R_{\alpha\beta}$ are estimated as indicated above. $R_{\beta\beta}$ is estimated with the correlation matrix for the regression of the β set on the α set using $R_{\alpha\alpha}$ and $R_{\alpha\beta}$, i.e.

$$(156a) \quad \hat{R}_{\beta\beta} = R_{\beta\alpha} R_{\alpha\alpha}^{-1} R_{\alpha\beta}$$

The loadings are then given as the eigenvectors of

$$(156b) \quad \hat{R} = \begin{bmatrix} R_{\alpha\alpha} & R_{\alpha\beta} \\ R_{\beta\alpha} & \hat{R}_{\beta\beta} \end{bmatrix}$$

Finally, the third method to be considered is a two step procedure. In the first step, the loadings and the corresponding component scores are estimated for the α set. In the second step, the loadings for the β set are estimated by taking the regression of the β set on the estimated components.

Let $\hat{\Gamma}_{\alpha}$ and $\hat{\Gamma}_{\beta}$ denote the estimated loadings for the two sets and let $\hat{\Lambda}$ be a diagonal matrix with the n largest eigenvalues of $R_{\alpha\alpha}$ as diagonal elements. We then obtain the loadings for the β set from

¹ It is possible that Horst has not seen these papers by Gibson (1960, 1961) as they are not included in the list of references of his book (Horst, 1965).

$$(157) \quad \hat{\Sigma}_{\beta} = R_{\beta\alpha} \hat{\Gamma}_{\alpha} \hat{\Lambda}^{-1}$$

Comments

As mentioned earlier, these two latter methods are applicable only when the observed variables can be divided in two sets, where one of the sets has no missing observations. Further, neither of them makes optimal use of the available data. This is partly reflected in the fact that the methods do not give consistent estimates of the loadings.

For the first of them we find, under regularity conditions, that

$$(158) \quad \text{plim}_{k \rightarrow \infty} \hat{R} = \begin{bmatrix} \Sigma_{\alpha\alpha} & \Sigma_{\alpha\beta} \\ \Sigma_{\beta\alpha} & \Sigma_{\beta\alpha} \Sigma_{\alpha\alpha}^{-1} \Sigma_{\alpha\beta} \end{bmatrix} = \hat{\Sigma}$$

But $\hat{\Sigma} \neq \Sigma$ and the eigenvalues and eigenvectors are also different. Thus, the method will give inconsistent estimates of the loadings.

The original form of the second method has the same property. Although the loadings for the α -set will be estimated consistently, the estimated loadings for the β -set will be inconsistent. We find that

$$(159a) \quad \text{plim}_{k \rightarrow \infty} \hat{\Gamma}_{\beta} = \Sigma_{\beta\alpha} \Gamma_{\alpha} \Lambda^{\star -1} = \Gamma_{\beta} \Lambda \Gamma_{\alpha} \Gamma_{\alpha}' \Lambda^{\star -1} \\ = \Gamma_{\beta} \Lambda \Lambda^{\star -1}$$

where Λ contains the factor variances and $\Lambda^{\star}$ the component variances. These will not be the same (Lyttkens, 1966). Thus

$$(159b) \quad \text{plim}_{k \rightarrow \infty} \hat{\Gamma}_{\beta} \neq \Gamma_{\beta}$$

However, it is possible to modify this latter method so that the estimates of Γ_{β} will be consistent. According to Lyttkens (1966) the

variance of a component equals the sum of the variance of the corresponding factor plus the variance of the residual. Thus

$$(159c) \quad \text{plim}_{k \rightarrow \infty} \hat{\Lambda} = \Lambda + \theta I$$

Then, letting $\hat{\theta}$ denote a consistent estimate of the residual variance, we obtain consistent estimates of the loadings for the β -set if we use

$$(160) \quad \hat{\hat{\Lambda}} = \hat{\Lambda} - \hat{\theta} I$$

instead of $\hat{\Lambda}$ in (157).

REFERENCES

- Afifi, A. and Elashoff, R.M. (1966): Missing Values in Multivariate Statistics I. Review of the Literature. *Journal of the American Statistical Association*, 61.
- Anderson, T.W. and Rubin, H. (1956): Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. V.
- Anderson, T.W. (1957): Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, 52.
- Anderson, T.W. (1958): *An Introduction to Multivariate Statistical Analysis*. Wiley & Sons, New York.
- Anderson, T.W. (1963): Asymptotic Theory for Principal Component Analysis. *The Annals of Mathematical Statistics*, 34.
- Bartlett, M.S. (1937): Some examples of statistical methods of research in agriculture. *Journal of the Royal Statistical Society, Supplement*, 4.
- Bartlett, M.S. (1950): Test of significance in factor analysis. *British Journal of Psychology, Statistical Section*, 3.
- Cramér, H. (1945): *Mathematical Methods of Statistics*. Almqvist & Wiksell, Uppsala.
- Christoffersson, A. (1970a): Analys av studietider i olika ämnen vid samhällsvetenskaplig fakultet. Läsåren 1964/65 och 1965/66. To be published in *Statistisk Tidskrift*.
- Christoffersson, A. (1970b): *The Multi-factor Model with Incomplete Data. Part I: Theoretical*. Stencil.
- Dalenius, T. (1957): *Sampling in Sweden*. Almqvist & Wiksell, Stockholm.
- Gibson, W.A. (1960): Remarks on Tucker's inter-battery method of factor analysis. *Psychometrika*, 25.
- Gibson, W.A. (1961): A retraction on inter-battery factor analysis. *Psychometrika*, 26.

- Gollob, H.F. (1968): A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrica*, 33.
- Fisher and Mackenzie (1923): Studies in crop variation. II. The manurial response of different potato varieties. *Journal of Agricultural Science*, 13.
- Hagnell, M. (1970): Personal communication.
- Hocking, R.R. and Smith, W.B. (1968): Estimation of Parameters in the Multivariate Normal Distribution with Missing Observations. *Journal of the American Statistical Association*, 63.
- Horst, P. (1965): *Factor Analysis of Data matrices*. Holt, Rinehart & Winston, INC., New York.
- Hotelling, H. (1933): Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24.
- Jöreskog, K.G. (1963): *Statistical Estimation in Factor Analysis*. Almqvist & Wiksell, Stockholm.
- Koort, P., Dalén, L. and Eriksson, G. (1961): *The Item Catalogue Method (ICM)*. Almqvist & Wiksell, Uppsala.
- Koort, P. and Barrefors, O. (1966): *Geografi 8. Lärarhandledning*, Provdelen. ILU, Uppsala.
- Koort, P. (1969): Personal communication.
- Lawley, D.N. (1940): The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, Section A*, 60.
- Lord, F.M. (1955): Estimation of parameters from incomplete data. *Journal of the American Statistical Association*, 50.
- Lytkens, E. (1966): On the Fixpoint Property of Wold's Iterative Estimation Method for Principal Components. Paper in *Multivariate Analysis*. Ed. Krishnaiah, P. Academic Press, New York.
- Mandel, J. (1961): Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association*, 56.

- National Central Bureau of Statistics, (SCB), Sweden (1967a): Length of studies in different subjects at the faculties of arts and sciences in 1964/65. Statistical Reports U 1967 :11.
- National Central Bureau of Statistics, (SCB), Sweden (1967b): Length of studies in different subjects at the faculties of arts and sciences in 1965/66. Statistical Reports U 1967:15.
- Pearson, K. (1901): On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 6.
- Rao, C.R. (1955): Estimations and tests of significance in factor analysis. *Psychometrika*, 29.
- Scheffé, H. (1954): *The Analysis of Variance*. Wiley & Sons, New York.
- Spearman, C. (1904): General intelligence, objectively determined and measured. *American Journal of Psychology*, 15.
- Tucker, L.R. (1958): An inter-battery method of factor analysis. *Psychometrika*, 23.
- Whittle, P. (1952): On Principal Components and Least Square Methods of Factor Analysis. *Skandinavisk Aktuarietidskrift*, 35.
- Wold, H. (1965): A fix-point theorem with econometric background: I – II. *Arkiv för matematik*, 6, nos. 12 – 13.
- Wold, H. (1956): Causal Inference from Observational Data. A Review of Ends and Means. *Journal of the Royal Statistical Society, Series A*, 119.
- Wold, H. (1966a): Nonlinear Estimation by Iterative Least Square Procedures. Paper in *Festschrift for J. Neyman*, Ed. David, F.N. Wiley & Sons, New York.
- Wold, H. (1966b): On Estimation of Principal Components and Related Models by Iterative Least Squares. Paper in *Multivariate Analysis*. Ed. Krishnaiah, P., Academic Press, New York.
- Yates, F. (1933): The analysis of replicated experiments when the field results are incomplete. *The Empire Journal of Experimental Agriculture*, 1.
- Young, G. (1941): Maximum Likelihood Estimation and Factor Analysis. *Psychometrika*, 6.

APPENDIX

TABLES FOR THE MONTE CARLO EXPERIMENTS

NOTATIONS:

σ^2 = theoretical residual variance.

σ_μ^2 = factor variance.

$\bar{\sigma}^2$ = arithmetic mean of the estimated residual variances.

S^2 = estimated variance of the estimated residual variances (based on 100 samples).

$D^2(\hat{\sigma}^2)$ = theoretical variance of the estimated residual variances.

ψ = theoretical covariance matrix for the loadings.¹

$\bar{\psi}$ = estimated covariance matrix for the loadings (based on 100 samples).¹

χ^2 = chi-square (with six degrees of freedom) obtained when testing the hypothesis that the covariance matrix for the loadings equals ψ .
The χ^2 which are "significant" at the 5% level are denoted with an ★.

Roman numerals, I, II, ... X, denote the different types of incompleteness.

In the tables the different columns are:

(1) = variable

(2) = % missing

(3) = arithmetic mean of the estimated loadings

(4) = length of the confidence interval based on the theoretical covariance matrix (see formula 51)

(5) = number of samples where the confidence interval, based on the theoretical covariance matrix does not cover the true loadings

(6) = number of samples where the confidence interval based on the estimated covariance matrix (see formula 84) does not cover the true loadings.

¹ As the matrix is symmetric, only the lower left part is given.

TABLE A

Loadings: $\gamma_1 = \frac{1}{2}$; $\gamma_2 = \frac{1}{2}$; $\gamma_3 = \frac{1}{2}$; $\gamma_4 = \frac{1}{2}$; $\sigma^2 = 4$; $\sigma_\mu^2 = 9$

100 replications

I

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.494	$\pm .19$	0	0
2	0	.497	$\pm .19$	1	0
3	0	.489	$\pm .19$	1	1
4	0	.500	$\pm .19$	2	3

$$\bar{\sigma}^2 = 4.01; S^2 = .1119$$

$$D^2(\hat{\sigma}^2) = .1077; \chi^2 = 11.18$$

$$10^3 \psi = \begin{bmatrix} 4.8 \\ -1.6 & 4.8 \\ -1.6 & -1.6 & 4.8 \\ -1.6 & -1.6 & -1.6 & 4.8 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 5.4 \\ -.7 & 3.7 \\ -1.7 & -1.5 & 4.2 \\ -3.0 & -1.6 & -.8 & 5.6 \end{bmatrix}$$

II

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.507	.21	0	0
2	10	.486	.20	2	2
3	10	.488	.21	1	0
4	10	.495	.20	2	2

$$\bar{\sigma}^2 = 4.03; S^2 = .1097$$

$$D^2(\hat{\sigma}^2) = .1245; \chi^2 = 7.72$$

$$10^3 \psi = \begin{bmatrix} 5.7 \\ -1.8 & 5.3 \\ -2.1 & -1.9 & 5.9 \\ -1.8 & -1.6 & -1.9 & 5.2 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 5.7 \\ -2.1 & 6.7 \\ -2.6 & -2.4 & 6.5 \\ -1.4 & -1.6 & -1.2 & 4.5 \end{bmatrix}$$

TABLE A cont.

III

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.492	$\pm .23$	0	0
2	20	.491	$\pm .23$	2	1
3	20	.489	$\pm .23$	0	0
4	20	.499	$\pm .23$	0	0

$$\sigma^2 = 3.98; S^2 = .1405$$

$$D^2(\hat{\sigma}^2) = .1475; \chi^2 = .86$$

$$10^3 \psi = \begin{bmatrix} 6.9 \\ -2.3 & 6.7 \\ -2.4 & -2.0 & 6.8 \\ -2.2 & -2.3 & -2.4 & 6.9 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 7.3 \\ -2.5 & 6.9 \\ -2.8 & -2.2 & 7.2 \\ -1.9 & -2.3 & -2.1 & 6.6 \end{bmatrix}$$

IV

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.477	$\pm .28$	1	1
2	30	.492	$\pm .25$	0	0
3	30	.499	$\pm .29$	2	2
4	30	.492	$\pm .27$	0	0

$$\bar{\sigma}^2 = 3.93; S^2 = .1853$$

$$D^2(\hat{\sigma}^2) = .1808; \chi^2 = 13.92^{\star}$$

$$10^3 \psi = \begin{bmatrix} 10.1 \\ -2.2 & 8.1 \\ -4.3 & -3.2 & 10.6 \\ -3.6 & -2.6 & -3.1 & 9.3 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 10.3 \\ -2.3 & 7.5 \\ -4.4 & -1.4 & 11.0 \\ -2.2 & -3.7 & -4.2 & 10.3 \end{bmatrix}$$

TABLE A cont.,

V

(1)	(2)	(3)	(4)	(5)	(6)
1	40	.485	$\pm .30$	0	1
2	40	.484	$\pm .27$	2	1
3	40	.491	$\pm .32$	0	2
4	40	.490	$\pm .28$	1	4

$$\bar{\sigma}^2 = 3.94; S^2 = .2878$$

$$D^2(\bar{\sigma}^2) = .2336; \chi^2 = 16.37^*$$

$$10^3 \psi = \begin{bmatrix} 11.2 \\ -2.7 & 9.3 \\ -5.0 & -4.0 & 12.8 \\ -3.5 & -2.6 & -3.8 & 9.9 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 9.7 \\ -3.5 & 13.3 \\ -2.6 & -5.3 & 12.7 \\ -3.1 & -4.8 & -4.0 & 12.5 \end{bmatrix}$$

VI

(1)	(2)	(3)	(4)	(5)	(6)
1	50	.481	$\pm .34$	1	2
2	50	.464	$\pm .36$	0	2
3	50	.468	$\pm .37$	1	4
4	50	.520	$\pm .44$	0	1

$$\bar{\sigma}^2 = 3.86; S^2 = .2487$$

$$D^2(\bar{\sigma}^2) = .3299; \chi^2 = 5.91$$

$$10^3 \psi = \begin{bmatrix} 14.5 \\ -4.2 & 17.0 \\ -3.2 & -4.6 & 17.1 \\ -7.0 & -8.1 & -9.3 & 24.4 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 15.3 \\ -4.1 & 13.9 \\ -1.8 & -3.3 & 16.4 \\ -5.7 & -5.8 & -8.7 & 17.9 \end{bmatrix}$$

TABLE A cont.

VII

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.500	$\pm .21$	0	1
2	10	.500	$\pm .22$	1	1
3	20	.491	$\pm .24$	0	0
4	30	.478	$\pm .26$	2	1

$$\bar{\sigma}^2 = 3.97; S^2 = .1121$$

$$D^2(\hat{\sigma}^2) = .1350; \chi^2 = 1.07$$

$$10^3 \psi = \begin{bmatrix} 5.9 \\ -1.2 & 6.1 \\ -2.0 & -2.1 & 7.3 \\ -2.7 & -2.8 & -3.2 & 8.7 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 5.8 \\ -1.3 & 6.1 \\ -1.9 & -1.5 & 6.9 \\ -3.0 & -3.2 & -3.6 & 11.0 \end{bmatrix}$$

VIII

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.493	$\pm .25$	0	2
2	20	.485	$\pm .24$	1	0
3	30	.498	$\pm .25$	1	1
4	40	.488	$\pm .26$	1	0

$$\bar{\sigma}^2 = 4.00; S^2 = .1457$$

$$D^2(\hat{\sigma}^2) = .1624; \chi^2 = 10.4$$

$$10^3 \psi = \begin{bmatrix} 8.0 \\ -2.0 & 7.2 \\ -2.8 & -2.6 & 8.4 \\ -3.2 & -2.6 & -3.0 & 8.8 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 9.5 \\ -1.3 & 7.3 \\ -2.6 & -3.9 & 8.6 \\ -5.7 & -2.3 & -1.4 & 9.8 \end{bmatrix}$$

TABLE A cont.

IX

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.466	$\pm .28$	2	1
2	30	.501	$\pm .29$	0	0
3	40	.507	$\pm .27$	1	1
4	50	.484	$\pm .31$	0	0

$$\bar{\sigma}^2 = 3.98; S^2 = .2040$$

$$D^2(\hat{\sigma}^2) = .2038; \chi^2 = 1.53$$

$$10^3 \psi = \begin{bmatrix} 9.7 \\ -2.9 & 11.0 \\ -2.5 & -3.3 & 9.3 \\ -4.2 & -4.8 & -3.5 & 12.6 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 9.2 \\ -3.3 & 10.4 \\ -3.1 & -2.7 & 9.5 \\ -2.5 & -4.2 & -3.7 & 11.6 \end{bmatrix}$$

X

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.487	$\pm .31$	1	1
2	40	.470	$\pm .33$	2	3
3	50	.483	$\pm .36$	2	2
4	60	.485	$\pm .41$	4	3

$$\bar{\sigma}^2 = 3.93; S^2 = .2408$$

$$D^2(\hat{\sigma}^2) = .2735; \chi^2 = 66.5^{\star}$$

$$10^3 \psi = \begin{bmatrix} 12.0 \\ -3.1 & 14.1 \\ -3.4 & -4.2 & 17.0 \\ -5.6 & -6.8 & -9.4 & 21.7 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 15.1 \\ 1.1 & 15.5 \\ -1.0 & 1.6 & 17.1 \\ -10.2 & -11.2 & -10.1 & 25.4 \end{bmatrix}$$

TABLE B

Loadings: $\gamma_1 = \frac{1}{2}$; $\gamma_2 = \frac{1}{2}$; $\gamma_3 = \frac{1}{2}$; $\gamma_4 = \frac{1}{2}$; $\sigma^2 = 4$; $\sigma_\mu^2 = 16$

100 replications

I

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.499	$\pm .13$	2	2
2	0	.497	$\pm .13$	1	1
3	0	.495	$\pm .13$	2	2
4	0	.498	$\pm .13$	1	1

$\bar{\sigma}^2 = 3.99$; $S^2 = .1061$
 $D^2(\hat{\sigma}^2) = .1077$; $\chi^2 = 3.31$

$$10^3 \psi = \begin{bmatrix} 2.3 & & & \\ -.8 & 2.3 & & \\ -.8 & -.8 & 2.3 & \\ -.8 & -.8 & -.8 & 2.3 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 2.5 & & & \\ -1.1 & 2.4 & & \\ -.8 & -.7 & 2.5 & \\ -.7 & -.6 & -.9 & 2.1 \end{bmatrix}$$

II

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.499	$\pm .14$	0	1
2	10	.499	$\pm .14$	1	1
3	10	.498	$\pm .14$	0	1
4	10	.492	$\pm .14$	2	2

$\bar{\sigma}^2 = 4.02$; $S^2 = .0794$
 $D^2(\hat{\sigma}^2) = .1245$; $\chi^2 = 16.60^*$

$$10^3 \psi = \begin{bmatrix} 2.5 & & & \\ -.8 & 2.5 & & \\ -.9 & -.9 & 2.6 & \\ -.8 & -.8 & -.8 & 2.5 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 2.6 & & & \\ -1.2 & 2.8 & & \\ -1.0 & -.6 & 3.1 & \\ -.5 & -1.0 & -1.6 & 3.1 \end{bmatrix}$$

TABLE B cont.

III

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.495	$\pm .17$	0	1
2	20	.495	$\pm .16$	0	0
3	20	.495	$\pm .17$	0	1
4	20	.501	$\pm .17$	2	2

$$\bar{\sigma}^2 = 3.96; S^2 = .1353$$

$$D^2(\bar{\sigma}^2) = .1475; \chi^2 = 9.84$$

$$10^3 \psi = \begin{bmatrix} 3.7 \\ -1.2 & 3.4 \\ -1.2 & -1.2 & 3.6 \\ -1.3 & -1.0 & -1.3 & 3.6 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.0 \\ -.8 & 2.4 \\ -1.3 & -.5 & 2.9 \\ -2.0 & -1.0 & -1.0 & 4.1 \end{bmatrix}$$

IV

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.502	$\pm .19$	0	1
2	30	.496	$\pm .17$	1	1
3	30	.494	$\pm .18$	0	0
4	30	.494	$\pm .19$	0	0

$$\bar{\sigma}^2 = 4.03; S^2 = .1923$$

$$D^2(\bar{\sigma}^2) = .1808; \chi^2 = 7.59$$

$$10^3 \psi = \begin{bmatrix} 4.4 \\ -1.0 & 3.8 \\ -1.6 & -1.3 & 4.1 \\ -1.8 & -1.4 & -1.2 & 4.4 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 3.5 \\ -.9 & 3.0 \\ -1.2 & -1.1 & 3.5 \\ -1.4 & -.9 & -1.3 & 3.5 \end{bmatrix}$$

TABLE B cont.

V	(1)	(2)	(3)	(4)	(5)	(6)	
	1	40	.496	$\pm .20$	0	0	$\bar{\sigma}^2 = 3.96; S^2 = .2386$
	2	40	.489	$\pm .22$	0	0	$D^2(\hat{\sigma}^2) = .2336; \chi^2 = 3.10$
	3	40	.489	$\pm .20$	0	0	
	4	40	.503	$\pm .27$	0	0	

$$10^3 \psi = \begin{bmatrix} 4.9 \\ -1.3 & 6.0 \\ -.7 & -1.4 & 5.1 \\ -2.9 & -3.3 & -2.9 & 9.1 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.1 \\ -1.1 & 5.8 \\ -.6 & -.9 & 4.5 \\ -2.3 & -3.4 & -3.0 & 8.7 \end{bmatrix}$$

VI	(1)	(2)	(3)	(4)	(5)	(6)	
	1	50	.498	$\pm .23$	1	1	$\bar{\sigma}^2 = 3.91; S^2 = .2762$
	2	50	.476	$\pm .25$	2	3	$D^2(\hat{\sigma}^2) = .3299; \chi^2 = 6.66$
	3	50	.490	$\pm .23$	0	1	
	4	50	.500	$\pm .30$	0	3	

$$10^3 \psi = \begin{bmatrix} 7.0 \\ -2.1 & 8.2 \\ -1.6 & -1.4 & 6.6 \\ -3.4 & -4.8 & -3.6 & 11.7 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 6.2 \\ -2.1 & 9.1 \\ -1.1 & -1.3 & 7.8 \\ -3.0 & -5.1 & -4.8 & 12.9 \end{bmatrix}$$

TABLE B cont.

VII

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.493	$\pm .14$	0	0
2	10	.492	$\pm .15$	1	0
3	20	.504	$\pm .16$	1	1
4	30	.499	$\pm .16$	0	0

$\bar{\sigma}^2 = 3.99$; $S^2 = .1037$
 $D^2(\bar{\sigma}^2) = .1350$; $\chi^2 = .53$

$$10^3 \psi = \begin{bmatrix} 2.6 \\ -.8 & 3.0 \\ -.9 & -1.1 & 3.2 \\ -1.0 & -1.1 & -1.2 & 3.3 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 2.6 \\ -.6 & 2.8 \\ -.8 & -1.1 & 3.1 \\ -1.0 & -1.1 & -1.2 & 3.3 \end{bmatrix}$$

VIII

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.488	$\pm .16$	1	2
2	20	.507	$\pm .17$	1	1
3	30	.495	$\pm .19$	1	0
4	40	.493	$\pm .19$	0	1

$\bar{\sigma}^2 = 3.99$; $S^2 = .1557$
 $D^2(\bar{\sigma}^2) = .1624$; $\chi^2 = 3.86$

$$10^3 \psi = \begin{bmatrix} 3.2 \\ -.7 & 3.9 \\ -1.2 & -1.6 & 4.8 \\ -1.3 & -1.6 & -2.0 & 4.8 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 3.9 \\ -.9 & 3.6 \\ -1.6 & -1.3 & 4.6 \\ -1.3 & -1.4 & -1.7 & 4.5 \end{bmatrix}$$

TABLE B cont.

IX

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.489	$\pm .18$	0	2
2	30	.497	$\pm .18$	0	1
3	40	.502	$\pm .19$	0	0
4	50	.494	$\pm .22$	0	0

$$\bar{\sigma}^2 = 4.00; S^2 = .1683$$

$$D^2(\hat{\sigma}^2) = .2038; \chi^2 = 5.68$$

$$10^3 \psi = \begin{bmatrix} 4.2 \\ - .9 & 4.3 \\ -1.3 & -1.1 & 4.5 \\ -2.0 & -2.3 & -2.1 & 6.3 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 3.4 \\ -1.1 & 4.5 \\ -1.3 & -1.4 & 4.3 \\ -1.0 & -2.2 & -1.4 & 4.9 \end{bmatrix}$$

X

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.493	$\pm .18$	0	1
2	40	.494	$\pm .21$	0	2
3	50	.501	$\pm .22$	0	0
4	60	.491	$\pm .25$	1	1

$$\bar{\sigma}^2 = 3.93; S^2 = .2596$$

$$D^2(\hat{\sigma}^2) = .2735; \chi^2 = 4.76$$

$$10^3 \psi = \begin{bmatrix} 4.3 \\ - .8 & 5.5 \\ -1.1 & -2.0 & 6.1 \\ -2.4 & -2.7 & -3.0 & 8.1 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.0 \\ -1.0 & 4.8 \\ - .5 & -1.5 & 4.8 \\ -2.4 & -2.3 & -2.7 & 7.5 \end{bmatrix}$$

TABLE C

Loadings: $\gamma_1 = \frac{1}{\sqrt{10}}$; $\gamma_2 = \frac{2}{\sqrt{10}}$; $\gamma_3 = \frac{-1}{\sqrt{10}}$; $\gamma_4 = \frac{-2}{\sqrt{10}}$; $\sigma^2 = 4$; $\sigma_\mu^2 = 9$

100 replications

I

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.314	$\pm .21$	1	2
2	0	.630	$\pm .17$	1	1
3	0	-.303	$\pm .21$	0	1
4	0	-.628	$\pm .17$	0	0

$$\bar{\sigma}^2 = 3.96; S^2 = .1009$$

$$D^2(\hat{\sigma}^2) = .1077; \chi^2 = 1.23$$

$$10^3 \psi = \begin{bmatrix} 5.8 \\ -1.3 & 3.9 \\ .6 & 1.3 & 5.8 \\ 1.3 & 2.6 & -1.3 & 3.9 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 6.2 \\ -1.1 & 3.5 \\ .9 & 1.4 & 6.0 \\ 1.4 & 2.2 & -1.2 & 3.5 \end{bmatrix}$$

II

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.304	$\pm .23$	1	0
2	10	.630	$\pm .18$	0	1
3	10	-.327	$\pm .23$	0	2
4	10	-.618	$\pm .19$	0	0

$$\bar{\sigma}^2 = 4.03; S^2 = .1170$$

$$D^2(\hat{\sigma}^2) = .1245; \chi^2 = 3.37$$

$$10^3 \psi = \begin{bmatrix} 6.5 \\ -1.3 & 4.1 \\ .7 & 1.4 & 6.6 \\ 1.6 & 2.8 & -1.6 & 4.4 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 6.4 \\ -2.2 & 4.5 \\ 1.1 & 1.6 & 6.9 \\ .3 & 2.5 & -1.5 & 3.4 \end{bmatrix}$$

TABLE C cont.

III

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.299	$\pm .25$	2	3
2	20	.608	$\pm .22$	1	2
3	20	-.301	$\pm .24$	2	1
4	20	-.647	$\pm .22$	1	1

$$\bar{\sigma}^2 = 3.99; S^2 = .1291$$

$$D^2(\hat{\sigma}^2) = .1475; \chi^2 = 9.47$$

$$10^3 \psi = \begin{bmatrix} 8.1 & & & \\ -1.8 & 6.1 & & \\ .7 & 1.7 & 7.4 & \\ 1.9 & 4.4 & -1.7 & 6.2 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 11.1 & & & \\ -2.6 & 6.7 & & \\ -.6 & 1.8 & 7.2 & \\ 2.7 & 4.5 & -1.3 & 6.5 \end{bmatrix}$$

IV

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.313	$\pm .27$	1	4
2	30	.628	$\pm .25$	0	1
3	30	-.330	$\pm .30$	1	1
4	30	-.598	$\pm .27$	4	3

$$\bar{\sigma}^2 = 4.01; S^2 = .1807$$

$$D^2(\hat{\sigma}^2) = .1808; \chi^2 = 10.42$$

$$10^3 \psi = \begin{bmatrix} 9.3 & & & \\ -1.3 & 8.0 & & \\ .4 & 2.3 & 12.3 & \\ 3.1 & 6.2 & -3.7 & 9.6 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 9.9 & & & \\ -1.2 & 8.3 & & \\ .6 & 4.7 & 13.5 & \\ 3.6 & 5.3 & -2.4 & 10.2 \end{bmatrix}$$

TABLE C cont.

V

(1)	(2)	(3)	(4)	(5)	(6)
1	40	.298	$\pm .37$	0	5
2	40	.615	$\pm .26$	0	2
3	40	-.302	$\pm .32$	0	1
4	40	-.622	$\pm .26$	2	2

$$\bar{\sigma}^2 = 3.96; S^2 = .2418$$

$$D^2(\hat{\sigma}^2) = .2236; \chi^2 = 9.52$$

$$10^3 \psi = \begin{bmatrix} 17.4 \\ -3.7 & 8.5 \\ 1.8 & 2.5 & 13.0 \\ 4.1 & 5.4 & -3.1 & 9.0 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 22.8 \\ -4.8 & 7.7 \\ 2.8 & 1.1 & 14.1 \\ 5.1 & 4.5 & -4.1 & 9.9 \end{bmatrix}$$

VI

(1)	(2)	(3)	(4)	(5)	(6)
1	50	.283	$\pm .40$	0	4
2	50	.603	$\pm .40$	2	3
3	50	-.318	$\pm .38$	0	3
4	50	-.599	$\pm .40$	3	5

$$\bar{\sigma}^2 = 3.91; S^2 = .2984$$

$$D^2(\hat{\sigma}^2) = .3299; \chi^2 = 16.37^{\star}$$

$$10^3 \psi = \begin{bmatrix} 20.2 \\ -6.8 & 20.6 \\ -.6 & 3.0 & 18.0 \\ 3.5 & 15.7 & -6.3 & 20.6 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 18.2 \\ -4.2 & 26.2 \\ -3.6 & 3.8 & 23.3 \\ 6.1 & 20.0 & -7.5 & 29.1 \end{bmatrix}$$

TABLE C cont.

VII

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.308	$\pm .23$	2	1
2	10	.626	$\pm .20$	2	1
3	20	-.313	$\pm .26$	1	2
4	30	-.622	$\pm .21$	0	2

$$\bar{\sigma}^2 = 4.02; S^2 = .1151$$

$$D^2(\hat{\sigma}^2) = .1350; \chi^2 = 2.31$$

$$10^3 \psi = \begin{bmatrix} 6.5 & & & \\ -1.2 & 5.2 & & \\ .9 & 1.6 & 8.8 & \\ 1.7 & 3.9 & -2.4 & 5.9 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 7.7 & & & \\ -1.3 & 4.9 & & \\ .9 & 1.6 & 9.4 & \\ 2.0 & 3.3 & -3.0 & 6.2 \end{bmatrix}$$

VIII

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.318	$\pm .24$	0	0
2	20	.617	$\pm .22$	1	0
3	30	-.310	$\pm .27$	1	1
4	40	-.629	$\pm .23$	1	1

$$\bar{\sigma}^2 = 3.97; S^2 = .1349$$

$$D^2(\hat{\sigma}^2) = .1624; \chi^2 = 10.96$$

$$10^3 \psi = \begin{bmatrix} 7.4 & & & \\ -1.3 & 6.4 & & \\ .8 & 1.9 & 9.1 & \\ 1.9 & 4.8 & -2.2 & 6.8 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 6.0 & & & \\ .0 & 5.0 & & \\ .2 & .9 & 8.3 & \\ 3.1 & 4.4 & -3.4 & 8.1 \end{bmatrix}$$

TABLE C cont.

IX

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.302	$\pm .28$	2	3
2	30	.610	$\pm .27$	2	5
3	40	-.299	$\pm .36$	2	2
4	50	-.627	$\pm .28$	2	3

$$\bar{\sigma}^2 = 3.98; S^2 = .1250$$

$$D^2(\hat{\sigma}^2) = .2038; \chi^2 = 15.68^{\star}$$

$$10^3 \psi = \begin{bmatrix} 10.0 \\ -1.9 & 9.2 \\ 1.5 & 3.4 & 16.1 \\ 2.4 & 6.5 & -3.9 & 9.7 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 10.9 \\ -3.8 & 13.3 \\ -1.4 & 4.8 & 12.0 \\ 1.7 & 8.8 & -4.3 & 12.3 \end{bmatrix}$$

X

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.312	$\pm .29$	1	1
2	40	.619	$\pm .27$	1	1
3	50	-.313	$\pm .36$	1	2
4	60	-.609	$\pm .30$	2	1

$$\bar{\sigma}^2 = 4.05; S^2 = .2757$$

$$D^2(\hat{\sigma}^2) = .2735; \chi^2 = 1.69$$

$$10^3 \psi = \begin{bmatrix} 10.6 \\ -1.5 & 9.4 \\ .8 & 3.1 & 16.7 \\ 3.4 & 7.1 & -4.8 & 11.2 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 11.7 \\ -1.1 & 9.5 \\ 1.3 & 3.8 & 16.9 \\ 3.1 & 7.1 & -5.1 & 12.9 \end{bmatrix}$$

TABLE D

Loadings: $\gamma_1 = \frac{1}{\sqrt{10}}$; $\gamma_2 = \frac{2}{\sqrt{10}}$; $\gamma_3 = \frac{-1}{\sqrt{10}}$; $\gamma_4 = \frac{-2}{\sqrt{10}}$; $\sigma^2 = 4$; $q_\mu^2 = 16$

100 replications

I

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.311	$\pm .15$	0	0
2	0	.624	$\pm .12$	0	0
3	0	-.315	$\pm .15$	1	2
4	0	-.636	$\pm .12$	1	0

$\bar{\sigma}^2 = 3.94$; $S^2 = .1006$
 $D^2(\hat{\sigma}^2) = .1077$; $\chi^2 = 4.40$

$$10^3 \psi = \begin{bmatrix} 2.8 \\ -.6 & 1.9 \\ .3 & .6 & 2.8 \\ .6 & 1.3 & -.6 & 1.9 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 2.9 \\ -.4 & 1.7 \\ .7 & .4 & 3.1 \\ .7 & 1.3 & -.8 & 2.1 \end{bmatrix}$$

II

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.311	$\pm .15$	1	2
2	10	.633	$\pm .13$	1	1
3	10	-.320	$\pm .16$	1	1
4	10	-.624	$\pm .13$	1	3

$\bar{\sigma}^2 = 4.02$; $S^2 = .1230$
 $D^2(\hat{\sigma}^2) = .1245$; $\chi^2 = 5.22$

$$10^3 \psi = \begin{bmatrix} 3.0 \\ -.7 & 2.2 \\ .3 & .7 & 3.1 \\ .7 & 1.5 & -.7 & 2.2 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 2.7 \\ -.8 & 2.6 \\ .0 & .7 & 3.5 \\ .6 & 1.8 & -1.0 & 2.6 \end{bmatrix}$$

TABLE D cont.

III

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.314	$\pm .18$	0	0
2	20	.634	$\pm .15$	2	1
3	20	-.316	$\pm .18$	0	0
4	20	-.620	$\pm .16$	1	0

$$\bar{\sigma}^2 = 3.97; S^2 = .1199$$

$$D^2(\hat{\sigma}^2) = .1475; \chi^2 = 0.30$$

$$10^3 \psi = \begin{bmatrix} 4.0 \\ -.8 & 2.9 \\ .4 & .8 & 4.0 \\ 1.0 & 2.1 & -1.0 & 3.1 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.2 \\ -.9 & 3.1 \\ .3 & .8 & 4.0 \\ .9 & 2.3 & -1.0 & 3.4 \end{bmatrix}$$

IV

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.324	$\pm .23$	0	0
2	30	.614	$\pm .18$	2	1
3	30	-.312	$\pm .18$	1	1
4	30	-.632	$\pm .17$	1	3

$$\bar{\sigma}^2 = 3.88; S^2 = .1231$$

$$D^2(\hat{\sigma}^2) = .1808; \chi^2 = 6.19$$

$$10^3 \psi = \begin{bmatrix} 6.9 \\ -1.9 & 4.1 \\ .6 & .8 & 4.3 \\ 1.3 & 2.8 & -1.0 & 3.9 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 6.8 \\ -1.9 & 4.3 \\ .1 & .5 & 4.9 \\ 1.6 & 3.0 & -1.7 & 4.7 \end{bmatrix}$$

TABLE D cont.

V

(1)	(2)	(3)	(4)	(5)	(6)
1	40	.299	$\pm .24$	2	1
2	40	.630	$\pm .18$	0	1
3	40	-.309	$\pm .22$	1	2
4	40	-.628	$\pm .20$	0	1

$$\bar{\sigma}^2 = 3.92; S^2 = .1767$$

$$D^2(\hat{\sigma}^2) = .2336; \chi^2 = 2.97$$

$$10^3 \psi = \begin{bmatrix} 7.4 & & & \\ -1.0 & 4.2 & & \\ .9 & 1.2 & 6.1 & \\ 2.3 & 3.2 & -1.4 & 5.0 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 7.9 & & & \\ -.3 & 3.8 & & \\ 1.5 & 1.4 & 7.1 & \\ 2.3 & 2.8 & -1.4 & 4.5 \end{bmatrix}$$

VI

(1)	(2)	(3)	(4)	(5)	(6)
1	50	.320	$\pm .33$	1	2
2	50	.619	$\pm .25$	1	1
3	50	-.302	$\pm .25$	0	1
4	50	-.624	$\pm .23$	1	1

$$\bar{\sigma}^2 = 3.99; S^2 = .3077$$

$$D^2(\hat{\sigma}^2) = .3299; \chi^2 = 8.55$$

$$10^3 \psi = \begin{bmatrix} 13.8 & & & \\ -2.9 & 7.8 & & \\ 2.0 & 2.6 & 7.7 & \\ 3.0 & 5.0 & -.2 & 6.6 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 12.3 & & & \\ -4.7 & 9.8 & & \\ 2.0 & 2.3 & 6.4 & \\ 1.0 & 5.1 & .4 & 5.1 \end{bmatrix}$$

TABLE D cont.

VII

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.326	$\pm .16$	1	1
2	10	.614	$\pm .16$	1	2
3	20	-.306	$\pm .19$	2	2
4	30	-.638	$\pm .16$	1	1

$$\bar{\sigma}^2 = 3.99; S^2 = .1015$$

$$D^2(\hat{\sigma}^2) = .1350; \chi^2 = 13.46^*$$

$$10^3 \psi = \begin{bmatrix} 3.4 \\ -.7 & 3.2 \\ .1 & .9 & 4.4 \\ 1.0 & 2.4 & -1.2 & 3.4 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.0 \\ -1.7 & 3.8 \\ .4 & .1 & 4.2 \\ .2 & 2.8 & -1.8 & 3.8 \end{bmatrix}$$

VIII

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.319	$\pm .17$	2	2
2	20	.630	$\pm .15$	0	0
3	30	-.321	$\pm .19$	0	0
4	40	-.619	$\pm .16$	0	2

$$\bar{\sigma}^2 = 4.02; S^2 = .1216$$

$$D^2(\hat{\sigma}^2) = .1624; \chi^2 = 11.9$$

$$10^3 \psi = \begin{bmatrix} 3.7 \\ -.6 & 2.8 \\ .3 & .9 & 4.8 \\ 1.1 & 2.1 & -1.4 & 3.4 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.6 \\ -.2 & 2.6 \\ -.1 & .2 & 3.8 \\ 2.1 & 2.3 & -1.7 & 4.1 \end{bmatrix}$$

TABLE D cont.

IX

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.306	$\pm .23$	1	2
2	30	.629	$\pm .20$	1	1
3	40	-.319	$\pm .21$	0	0
4	50	-.622	$\pm .26$	0	0

$$\bar{\sigma}^2 = 4.00; S^2 = .1469$$

$$D^2(\hat{\sigma}^2) = .2038; \chi^2 = 8.84$$

$$10^3 \psi = \begin{bmatrix} 6.6 & & & \\ .0 & 5.4 & & \\ .2 & -.2 & 5.7 & \\ 3.1 & 5.4 & -2.9 & 8.5 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 6.2 & & & \\ -.1 & 4.9 & & \\ -.5 & .9 & 5.2 & \\ 2.0 & 4.2 & -2.0 & 6.5 \end{bmatrix}$$

X

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.306	$\pm .24$	1	1
2	40	.622	$\pm .24$	0	2
3	50	-.300	$\pm .32$	0	1
4	60	-.626	$\pm .29$	3	2

$$\bar{\sigma}^2 = 3.97; S^2 = .2707$$

$$D^2(\hat{\sigma}^2) = .2735; \chi^2 = 3.57$$

$$10^3 \psi = \begin{bmatrix} 7.3 & & & \\ .6 & 7.2 & & \\ 1.6 & 1.8 & 13.0 & \\ 3.4 & 6.7 & -3.9 & 10.4 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 6.3 & & & \\ .1 & 7.8 & & \\ .7 & 1.8 & 12.3 & \\ 3.2 & 6.9 & -3.5 & 10.9 \end{bmatrix}$$

TABLE E

Loadings: $\gamma_1 = \frac{1}{\sqrt{10}}$; $\gamma_2 = \frac{2}{\sqrt{10}}$; $\gamma_3 = \frac{-1}{\sqrt{10}}$; $\gamma_4 = \frac{-2}{\sqrt{10}}$; $\sigma^2 = 4$; $\sigma_\mu^2 = 16$

100 replications

I

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.311	$\pm .15$	0	0
2	0	.624	$\pm .12$	0	0
3	0	-.315	$\pm .15$	1	2
4	0	-.636	$\pm .12$	1	0

$$\bar{\sigma}^2 = 3.90; S^2 = .0979$$

$$D^2(\hat{\sigma}^2) = .1077; \chi^2 = 4.36$$

$$10^3 \psi = \begin{bmatrix} 2.8 & & & \\ -.6 & 1.9 & & \\ .3 & .6 & 2.8 & \\ .6 & 1.3 & -.6 & 1.9 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 3.0 & & & \\ -.4 & 1.7 & & \\ .7 & .4 & 3.2 & \\ .7 & 1.3 & -.8 & 2.0 \end{bmatrix}$$

II

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.312	$\pm .15$	1	2
2	10	.632	$\pm .13$	1	1
3	10	-.321	$\pm .16$	1	1
4	10	-.623	$\pm .13$	1	3

$$\bar{\sigma}^2 = 3.96; S^2 = .1244$$

$$D^2(\hat{\sigma}^2) = .1245; \chi^2 = 5.00$$

$$10^3 \psi = \begin{bmatrix} 3.0 & & & \\ -.7 & 2.2 & & \\ .3 & .7 & 3.1 & \\ .7 & 1.5 & -.7 & 2.2 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 2.8 & & & \\ -.8 & 2.6 & & \\ .0 & .6 & 3.5 & \\ .6 & 1.8 & -1.1 & 2.6 \end{bmatrix}$$

TABLE E cont.

III

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.314	$\pm .18$	0	0
2	20	.633	$\pm .15$	2	1
3	20	-.316	$\pm .18$	0	1
4	20	-.622	$\pm .16$	1	0

$$\bar{\sigma}^2 = 3.92; S^2 = .1157$$

$$D^2(\hat{\sigma}^2) = .1475; \chi^2 = .93$$

$$10^3 \psi = \begin{bmatrix} 4.0 & & & \\ - .8 & 2.9 & & \\ .4 & .8 & 4.0 & \\ 1.0 & 2.1 & -1.0 & 3.1 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.2 & & & \\ - .9 & 3.1 & & \\ .3 & .7 & 4.1 & \\ 1.0 & 2.4 & -1.2 & 3.6 \end{bmatrix}$$

IV

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.322	$\pm .23$	0	0
2	30	.612	$\pm .18$	1	1
3	30	-.314	$\pm .18$	1	0
4	30	-.634	$\pm .17$	1	3

$$\bar{\sigma}^2 = 3.85; S^2 = .1254$$

$$D^2(\hat{\sigma}^2) = .1808; \chi^2 = 6.15$$

$$10^3 \psi = \begin{bmatrix} 6.9 & & & \\ -1.9 & 4.1 & & \\ .6 & .8 & 4.3 & \\ 1.3 & 2.8 & -1.0 & 3.9 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 6.9 & & & \\ -1.7 & 4.1 & & \\ .2 & .4 & 5.0 & \\ 1.7 & 2.9 & -1.8 & 4.7 \end{bmatrix}$$

TABLE E cont.

V

(1)	(2)	(3)	(4)	(5)	(6)
1	40	.300	$\pm .24$	2	1
2	40	.631	$\pm .18$	0	1
3	40	-.307	$\pm .22$	1	2
4	40	-.628	$\pm .20$	0	1

$$\bar{\sigma}^2 = 3.87; S^2 = .1742$$

$$D^2(\hat{\sigma}^2) = .2336; \chi^2 = 3.68$$

$$10^3 \psi = \begin{bmatrix} 7.4 & & & \\ -1.0 & 4.2 & & \\ .9 & 1.2 & 6.1 & \\ 2.3 & 3.2 & -1.4 & 5.0 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 8.0 & & & \\ - .3 & 3.8 & & \\ 1.5 & 1.5 & 7.5 & \\ 2.3 & 2.8 & -1.4 & 4.6 \end{bmatrix}$$

VI

(1)	(2)	(3)	(4)	(5)	(6)
1	50	.311	$\pm .33$	1	1
2	50	.615	$\pm .25$	1	1
3	50	-.303	$\pm .25$	0	0
4	50	-.632	$\pm .23$	1	1

$$\bar{\sigma}^2 = 3.92; S^2 = .2904$$

$$D^2(\hat{\sigma}^2) = .3299; \chi^2 = 6.69$$

$$10^3 \psi = \begin{bmatrix} 13.8 & & & \\ -2.9 & 7.8 & & \\ 2.0 & 2.6 & 7.7 & \\ 3.0 & 5.0 & -.2 & 6.6 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 12.9 & & & \\ -4.4 & 9.3 & & \\ 1.9 & 2.5 & 6.4 & \\ 1.4 & 4.8 & .6 & 5.0 \end{bmatrix}$$

TABLE E cont.

VII

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.326	$\pm .16$	1	1
2	10	.616	$\pm .16$	1	2
3	20	-.307	$\pm .19$	2	2
4	30	-.636	$\pm .16$	1	1

$$\bar{\sigma}^2 = 3.95; S^2 = .0997$$

$$D^2(\hat{\sigma}^2) = .1350; \chi^2 = 14.39$$

$$10^3 \psi = \begin{bmatrix} 3.4 & & & \\ - .7 & 3.2 & & \\ .1 & .9 & 4.4 & \\ 1.0 & 2.4 & -1.2 & 3.4 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.0 & & & \\ -1.7 & 3.8 & & \\ .3 & .1 & 4.2 & \\ .2 & 2.8 & -1.9 & 3.9 \end{bmatrix}$$

VIII

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.321	$\pm .17$	1	1
2	20	.630	$\pm .15$	0	0
3	30	-.321	$\pm .19$	0	0
4	40	-.618	$\pm .16$	0	2

$$\bar{\sigma}^2 = 3.97; S^2 = .1177$$

$$D^2(\hat{\sigma}^2) = .1624; \chi^2 = 11.63$$

$$10^3 \psi = \begin{bmatrix} 3.7 & & & \\ - .6 & 2.8 & & \\ .3 & .9 & 4.8 & \\ 1.1 & 2.1 & -1.4 & 3.3 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.6 & & & \\ - .3 & 2.6 & & \\ - .1 & .2 & 3.7 & \\ 2.1 & 2.2 & -1.6 & 4.0 \end{bmatrix}$$

TABLE E cont.

IX

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.306	$\pm .23$	2	2
2	30	.633	$\pm .21$	1	1
3	40	-.321	$\pm .21$	0	0
4	50	-.616	$\pm .26$	0	0

$$\bar{\sigma}^2 = 3.94; S^2 = .1444$$

$$D^2(\hat{\sigma}^2) = .2034; \chi^2 = 9.54$$

$$10^3 \psi = \begin{bmatrix} 6.6 & & & \\ .0 & 5.4 & & \\ .2 & -.2 & 5.7 & \\ 3.1 & 5.4 & -2.9 & 8.5 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 6.1 & & & \\ -.8 & 5.1 & & \\ -.4 & 1.0 & 5.1 & \\ 2.1 & 4.3 & -1.9 & 6.6 \end{bmatrix}$$

X

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.314	$\pm .24$	2	2
2	40	.648	$\pm .24$	0	3
3	50	-.313	$\pm .32$	0	0
4	60	-.583	$\pm .29$	4	3

$$\bar{\sigma}^2 = 3.96; S^2 = .2741$$

$$D^2(\hat{\sigma}^2) = .2735; \chi^2 = 5.03$$

$$10^3 \psi = \begin{bmatrix} 7.3 & & & \\ .6 & 7.2 & & \\ 1.6 & 1.8 & 13.0 & \\ 3.4 & 6.7 & -3.9 & 10.4 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 7.0 & & & \\ .5 & 8.4 & & \\ .6 & 2.2 & 14.3 & \\ 3.9 & 7.9 & -5.0 & 13.9 \end{bmatrix}$$

TABLE F

Loadings: $\gamma_1 = \frac{1}{\sqrt{10}}$; $\gamma_2 = \frac{2}{\sqrt{10}}$; $\gamma_3 = \frac{-1}{\sqrt{10}}$; $\gamma_4 = \frac{-2}{\sqrt{10}}$; $\sigma^2 = 4$; $\sigma_\mu^2 = 16$
100 replications

I	(1)	(2)	(3)	(4)	(5)	(6)	
	1	0	.320	$\pm .15$	1	0	$\bar{\sigma}^2 = 4.02$; $S^2 = .0952$
	2	0	.637	$\pm .12$	0	0	$D^2(\hat{\sigma}^2) = .1077$; $\chi^2 = 7.59$
	3	0	-.312	$\pm .15$	3	0	
	4	0	-.620	$\pm .12$	2	1	

$$10^3 \psi = \begin{bmatrix} 2.8 & & & \\ - .6 & 1.9 & & \\ .3 & .6 & 2.8 & \\ .6 & 1.3 & - .6 & 1.9 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 3.0 & & & \\ - .5 & 1.9 & & \\ .1 & .9 & 3.8 & \\ 1.0 & 1.3 & - .9 & 2.3 \end{bmatrix}$$

II	(1)	(2)	(3)	(4)	(5)	(6)	
	1	10	.312	$\pm .16$	0	0	$\bar{\sigma}^2 = 3.98$; $S^2 = 1326$
	2	10	.635	$\pm .13$	0	0	$D^2(\hat{\sigma}^2) = .1245$; $\chi^2 = 10.39$
	3	10	-.308	$\pm .16$	0	0	
	4	10	-.629	$\pm .13$	2	0	

$$10^3 \psi = \begin{bmatrix} 3.3 & & & \\ - .7 & 2.2 & & \\ .4 & .7 & 3.3 & \\ .7 & 1.5 & - .8 & 2.2 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 2.7 & & & \\ - .3 & 1.5 & & \\ .6 & .4 & 3.2 & \\ .7 & 1.2 & - .9 & 2.0 \end{bmatrix}$$

TABLE F cont.

III

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.311	$\pm .17$	3	2
2	20	.629	$\pm .15$	0	0
3	20	-.317	$\pm .17$	2	2
4	20	-.628	$\pm .15$	2	0

$$\bar{\sigma}^2 = 3.98; S^2 = .1457$$

$$D^2(\hat{\sigma}^2) = .1475; \chi^2 = 17.14^{\star}$$

$$10^3 \psi = \begin{bmatrix} 3.9 \\ - .9 & 2.8 \\ .4 & .9 & 3.9 \\ .9 & 1.9 & -.9 & 2.8 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.4 \\ -.8 & 2.0 \\ .5 & .8 & 5.2 \\ 1.1 & 1.3 & -1.5 & 2.7 \end{bmatrix}$$

IV

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.328	$\pm .19$	2	1
2	30	.622	$\pm .17$	0	3
3	30	-.314	$\pm .19$	1	1
4	30	-.622	$\pm .17$	2	0

$$\bar{\sigma}^2 = 3.94; S^2 = .1769$$

$$D^2(\hat{\sigma}^2) = .1808; \chi^2 = 7.50$$

$$10^3 \psi = \begin{bmatrix} 4.7 \\ -1.0 & 3.7 \\ .5 & 1.0 & 4.7 \\ 1.1 & 2.7 & -1.1 & 3.7 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 5.7 \\ -1.1 & 4.4 \\ .5 & 1.2 & 5.4 \\ 1.4 & 3.2 & -1.1 & 4.7 \end{bmatrix}$$

TABLE F cont.

V

(1)	(2)	(3)	(4)	(5)	(6)
1	40	.318	$\pm .22$	0	1
2	40	.616	$\pm .20$	2	2
3	40	-.307	$\pm .22$	1	1
4	40	-.634	$\pm .20$	1	0

$\bar{\sigma}^2 = 3.97$; $S^2 = .2080$
 $D^2(\hat{\sigma}^2) = .2336$; $\chi^2 = 6.11$

$$10^3 \psi = \begin{bmatrix} 6.4 \\ -1.4 & 5.0 \\ .6 & 1.4 & 6.2 \\ 1.5 & 3.5 & -1.3 & 4.9 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 7.0 \\ -2.2 & 6.0 \\ -.3 & 1.1 & 5.2 \\ 1.5 & 4.2 & -1.4 & 5.8 \end{bmatrix}$$

VI

(1)	(2)	(3)	(4)	(5)	(6)
1	50	.313	$\pm .27$	3	2
2	50	.617	$\pm .23$	2	1
3	50	-.310	$\pm .26$	2	1
4	50	-.620	$\pm .24$	2	1

$\bar{\sigma}^2 = 4.02$; $S^2 = .2755$
 $D^2(\hat{\sigma}^2) = .3299$; $\chi^2 = 17.23$

$$10^3 \psi = \begin{bmatrix} 9.4 \\ -1.8 & 6.8 \\ .8 & 1.8 & 8.9 \\ 2.5 & 5.0 & -2.2 & 7.3 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 9.6 \\ -1.2 & 9.9 \\ 2.2 & 1.8 & 11.1 \\ 2.3 & 7.7 & -2.9 & 10.8 \end{bmatrix}$$

TABLE F cont.

VII

(1)	(2)	(3)	(4)	(5)	(6)
1	0	.314	$\pm .16$	0	0
2	10	.631	$\pm .14$	0	0
3	20	-.314	$\pm .17$	0	1
4	30	-.626	$\pm .15$	1	0

$$\bar{\sigma}^2 = 3.88; S^2 = .1069$$

$$D^2(\hat{\sigma}^2) = .1350; \chi^2 = 6.90$$

$$10^3 \psi = \begin{bmatrix} 3.2 \\ -.6 & 2.5 \\ .3 & .7 & 3.9 \\ .9 & 1.9 & -1.0 & 2.8 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 3.4 \\ -.7 & 2.7 \\ .2 & 1.4 & 4.2 \\ .9 & 1.6 & -.6 & 2.5 \end{bmatrix}$$

VIII

(1)	(2)	(3)	(4)	(5)	(6)
1	10	.312	$\pm .17$	0	1
2	20	.632	$\pm .16$	1	0
3	30	-.305	$\pm .19$	1	0
4	40	-.629	$\pm .17$	0	0

$$\bar{\sigma}^2 = 3.99; S^2 = .1464$$

$$D^2(\hat{\sigma}^2) = .1624; \chi^2 = 6.20$$

$$10^3 \psi = \begin{bmatrix} 3.8 \\ -.6 & 3.2 \\ .3 & .9 & 4.7 \\ 1.1 & 2.5 & -1.3 & 3.7 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.4 \\ -.3 & 2.7 \\ 1.0 & 1.0 & 5.0 \\ 1.5 & 2.1 & -.9 & 3.4 \end{bmatrix}$$

TABLE F cont.

IX

(1)	(2)	(3)	(4)	(5)	(6)
1	20	.310	$\pm .19$	0	1
2	30	.630	$\pm .18$	2	2
3	40	-.306	$\pm .22$	1	3
4	50	-.625	$\pm .20$	2	2

$$\bar{\sigma}^2 = 3.95; S^2 = .1659$$

$$D^2(\bar{\sigma}^2) = .2038; \chi^2 = 18.01 \star$$

$$10^3 \psi = \begin{bmatrix} 4.6 \\ - .7 & 4.3 \\ .4 & 1.0 & 6.1 \\ 1.4 & 3.5 & -1.9 & 5.1 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 4.1 \\ - .7 & 4.6 \\ .4 & - .3 & 7.9 \\ .9 & 4.1 & -3.1 & 5.8 \end{bmatrix}$$

X

(1)	(2)	(3)	(4)	(5)	(6)
1	30	.305	$\pm .22$	2	1
2	40	.635	$\pm .22$	1	4
3	50	-.288	$\pm .26$	1	5
4	60	-.621	$\pm .25$	2	2

$$\bar{\sigma}^2 = 3.97; S^2 = .2387$$

$$D^2(\bar{\sigma}^2) = .2735; \chi^2 = 11.77$$

$$10^3 \psi = \begin{bmatrix} 6.3 \\ - .7 & 6.0 \\ .3 & 1.2 & 8.7 \\ 2.3 & 5.1 & -3.0 & 7.2 \end{bmatrix}$$

$$10^3 \bar{\psi} = \begin{bmatrix} 8.0 \\ -1.4 & 7.5 \\ - .5 & 2.9 & 10.7 \\ 3.3 & 4.8 & -2.7 & 8.4 \end{bmatrix}$$

